Robust Signal Processing for Dependent Data with Applications in Biomedicine

TECHNISCHE UNIVERSITÄT DARMSTADT

Michael Muma

SIGNAL PROCESSING GROUP

Signal Processing Group
Technische Universität Darmstadt
muma@spg.tu-darmstadt.de

Outline

Introduction

Real Data Example

Motion Artifacts in Intracranial Pressure Signal

Intracranial Pressure (ICP)

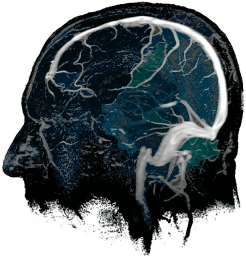
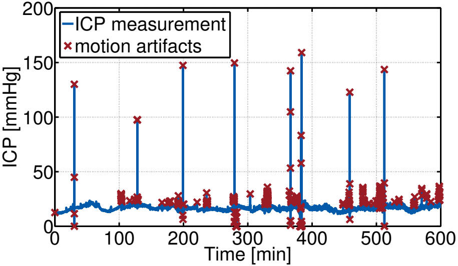


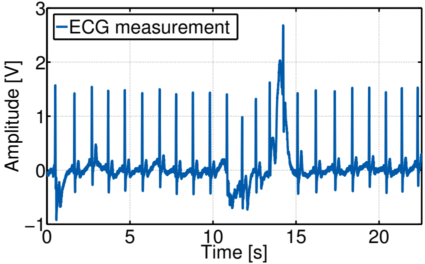
photo: ESA



ten hour excerpt of a typical ICP measurement

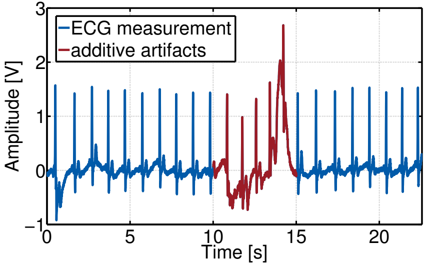
artifacts → robust ICP forecasting → early interventions for patients with traumatic brain injuries

Electrocardiogram (ECG)



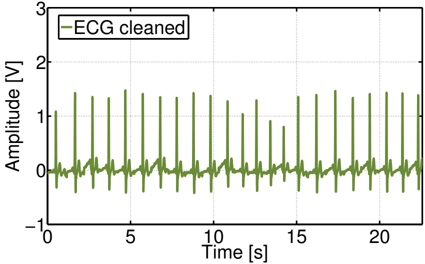
23 second excerpt of ECG from a psychological study

Electrocardiogram (ECG)



23 second excerpt of ECG from a psychological study

Electrocardiogram (ECG)

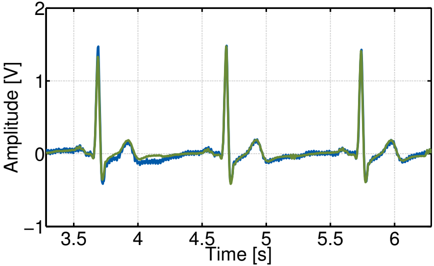


23 second excerpt of ECG from a psychological study

motion artifact cancellation → reliable ECG analysis

TECHNISCHE UNIVERSITÄT DARMSTADT

Electrocardiogram (ECG)



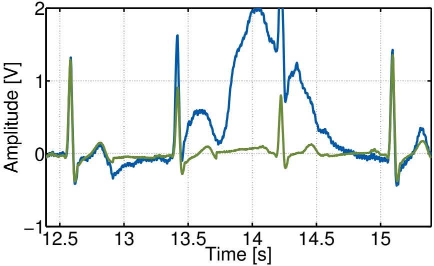
detail: artifact-free segment

Real Data Example

Electrocardiogram (ECG)

TECHNISCHE UNIVERSITÄT DARMSTADT

Electrocardiogram (ECG)



detail: artifact-contaminated segment

Real Data Example

Electrocardiogram (ECG)

TECHNISCHE UNIVERSITÄT DARMSTADT

Reminder on Aims of Robust Methods

- *‘The procedure should behave “reasonably well” at the assumed model.’*
- *‘The effect of an erroneous observation, even if it takes an arbitrary value, should not have a large impact on the method.’*
- *’Somewhat larger deviations from the model should not cause a catastrophe.’*

Real Data Example

Electrocardiogram (ECG)

TECHNISCHE UNIVERSITÄT DARMSTADT

Intuitive Example: Nonparametric Spectral Analysis of ECG [9]

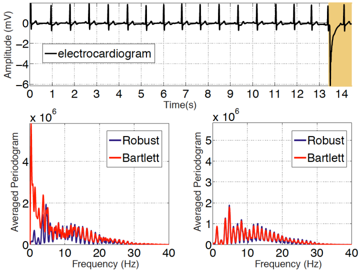
Bartlett estimator

- Split the measurement $\mathbf{x}$ into M parts $\mathbf{x}_m$ of the same length
- Compute the periodogram $I_{XX}(e^{j\omega}, m)$ for each $\mathbf{x}_m$
- $\hat{C}_{XX}(e^{j\omega}) = \dfrac{1}{M}\sum_{m=1}^{M} I_{XX}(e^{j\omega}, m)$

Robust Bartlett estimator

- Replace the sample mean in Bartlett's estimator by a robust location estimate (e.g. M-estimate or sample median)

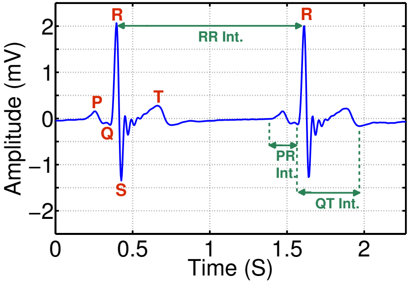
Intuitive Example: Nonparametric Spectral Analysis of ECG [9]



- Similar results for clean measurements (right)
- Robust estimator maintains performace in presence of artefacts (left)
  - ⊕ Simple and computationally cheap robust alternative
  - ⊖ Can only be expected to work well, when majority of parts $\mathbf{x}_m$, $m = 1, \ldots, M$ does not contain outliers

Example: Outlier Cleaning for RR Traces [57]

- important, e.g. for heart rate variability analysis, arrhythmia detection, fitness monitoring
- can be derived, from ECG by detecting the peaks



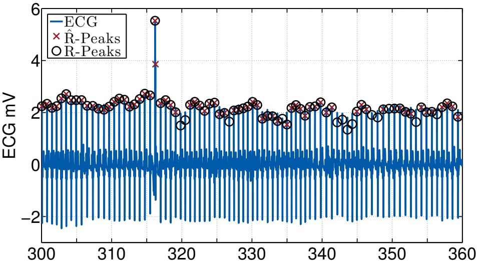
Real Data Example

RR Trace

TECHNISCHE UNIVERSITÄT DARMSTADT

Example: Outlier Cleaning for RR Traces [57]

- important, e.g. for heart rate variability analysis, arrhythmia detection, fitness monitoring
- can be derived, from ECG by detecting the peaks

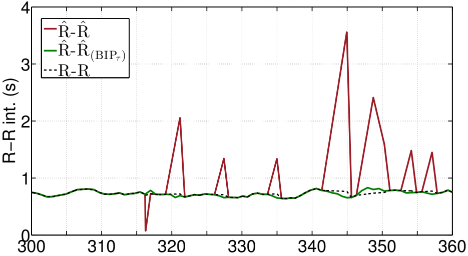


Real Data Example

RRTrace

Example: Outlier Cleaning for RR Traces [57]

- important, e.g. for heart rate variability analysis, arrhythmia detection, fitness monitoring
- can be derived, from ECG by detecting the peaks



true R-R series (black); estimated R-R series (red); outlier cleaned R-R series (green)

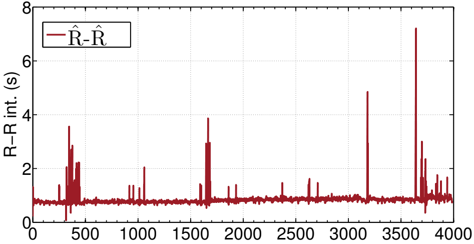
- missed detections and false alarms $\rightarrow$ outliers in R-R series

Real Data Example

RR Trace

Example: Outlier Cleaning for RR Traces [57]

- important, e.g. for heart rate variability analysis, arrhythmia detection, fitness monitoring
- can be derived, from ECG by detecting the peaks



complete estimated R-R series

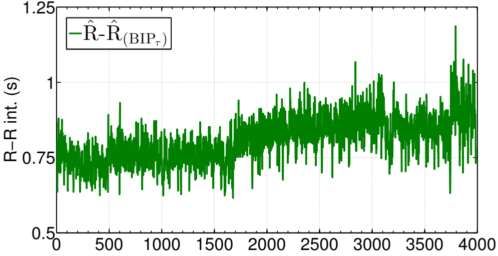
- missed detections and false alarms $\rightarrow$ outliers in R-R series

Real Data Example

RR Trace

Example: Outlier Cleaning for RR Traces [57]

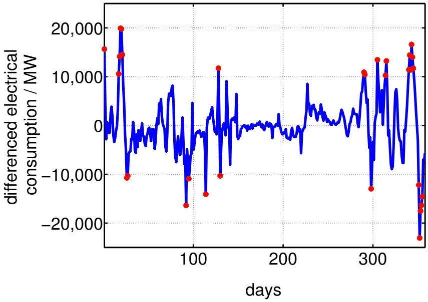
- important, e.g. for heart rate variability analysis, arrhythmia detection, fitness monitoring
- can be derived, from ECG by detecting the peaks



R-R series after outlier cleaning based on robust parameter estimation

Further Applications

- short-term load forecasting



- econometrics
- audio restoration
- ...

Signal and Outlier Models

Introduction

Basic Assumptions

(Local) Stationarity

- approximately holds for some signals, e.g. local stationarity for speech $< 15 - 30$ ms
- holds after suitable pre-processing for many others, e.g. differentiation, empirical mode decomposition

Most Popular Model

- As in classical signal processing for dependent data: AutoRegressive Moving Average (ARMA)

Autoregressive Moving Average (ARMA) Model

TECHNISCHE UNIVERSITÄT DARMSTADT

Autoregressive Moving Average ARMA(p,q)

$$y_t = \mu + \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu) + a_t - \sum_{i=1}^{q} \theta_i a_{t-i} \tag{1}$$

- y_t: observations with mean value μ
- a_t: i.i.d. Gaussian random variables (“innovations”) with finite variance
- $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$: autoregressive parameters
- $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$: moving average parameters
- $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$: parameter vector of ARMA

Autoregressive Moving Average (ARMA) Model

TECHNISCHE UNIVERSITÄT DARMSTADT

Let

$$\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i \quad \text{and} \quad \theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i.$$

then

$$a_t^e(\boldsymbol{\beta}) = \theta^{-1}(B)\phi(B)(y_t - \mu),$$

Assume

- all roots of $\phi(B)$ and $\theta(B)$ are outside the unit circle
  $\rightarrow$ stationary and invertible models
- $\phi(B)$ and $\theta(B)$ do not have common roots
  $\rightarrow$ identifiable models

Autoregressive Moving Average (ARMA) Model

TECHNISCHE UNIVERSITÄT DARMSTADT

Recursion for innovations

$$a_t^e(\boldsymbol{\beta}) = y_t - \mu - \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu) + \sum_{i=1}^{q} \theta_i a_{t-i}^e(\boldsymbol{\beta}), \quad t = p+1, p+2, \ldots$$

- $a_t^e(\boldsymbol{\beta}) = a_t$.

ARMA Parameter Estimation

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad f(a_t^e(\boldsymbol{\beta}))$$

- $\mathbf{a}_n(\boldsymbol{\beta}) = (a_1^e(\boldsymbol{\beta}), a_2^e(\boldsymbol{\beta}), \ldots, a_n^e(\boldsymbol{\beta}))$

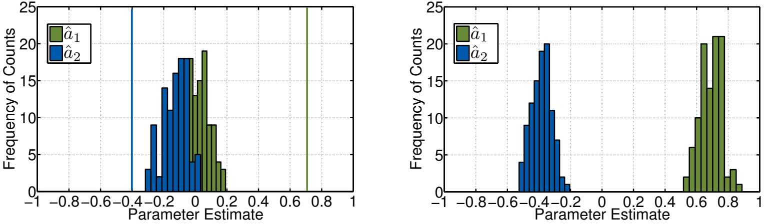
Autoregressive Moving Average (ARMA) Model

TECHNISCHE UNIVERSITÄT DARMSTADT

Robust Parameter Estimation

- estimate the ARMA parameters reliably given a finite number of (partially corrupted) observations

Example: AR(2), $\phi = (0.7, -0.4)$, $n = 200$, 10 % additive outliers



left: classical estimator, right: robust estimator

Autoregressive Moving Average (ARMA) Model

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Robust Model Order Selection

- estimate p and q: select the candidate model that minimizes

$$\mathrm{IC}(p, q) = \textbf{robust data fit} + \textbf{model complexity penalty}$$

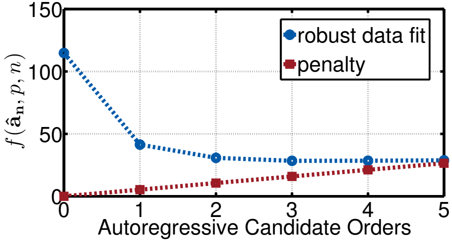
Autoregressive Moving Average (ARMA) Model

Robust Model Order Selection

- estimate p and q: select the candidate model that minimizes

$$\mathrm{IC}(p, q) = \textbf{robust data fit} + \textbf{model complexity penalty}$$

Example: AR(2), $\phi = (0.7, -0.4)$, $n = 200$, 10 % additive outliers



Autoregressive Moving Average (ARMA) Model

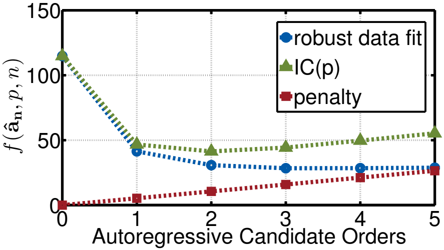
TECHNISCHE UNIVERSITÄT DARMSTADT

Robust Model Order Selection

- estimate p and q: select the candidate model that minimizes

$$\mathrm{IC}(p, q) = \textbf{robust data fit} + \textbf{model complexity penalty}$$

Example: AR(2), $\phi = (0.7, -0.4)$, $n = 200$, 10 % additive outliers



Autoregressive Moving Average (ARMA) Model

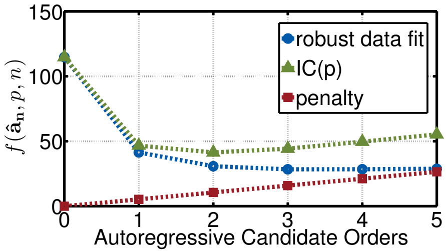
TECHNISCHE UNIVERSITÄT DARMSTADT

Robust Model Order Selection

- estimate p and q: select the candidate model that minimizes

$$\text{IC}(p,q) = \underbrace{\textbf{robust data fit}}_{f(a_t^e(\hat{\beta}(p,q)))} + \underbrace{\textbf{model complexity penalty}}_{g(p,q,n)}$$

Example: AR(2), $\phi = (0.7, -0.4)$, $n = 200$, 10 % additive outliers



Outlier Models

Additive Outliers

TECHNISCHE UNIVERSITÄT DARMSTADT

Additive Outliers (AO)

$$y_t^{\varepsilon} = x_t + \xi_t^{\varepsilon} w_t,$$

- y_t^{ε}: contaminated observations
- x_t: core process which follows Eq. (1)
- w_t: contaminating process, independent of x_t
- ξ_t^{ε}: stationary random process

$$\xi_t^{\varepsilon} = \begin{cases} 1 & \text{with probability } \varepsilon \\ 0 & \text{with probability } (1-\varepsilon). \end{cases}$$

Outlier Models

Replacement Outliers

TECHNISCHE UNIVERSITÄT DARMSTADT

Replacement Outliers (RO)

$$y_t^{\varepsilon} = (1 - \xi_t^{\varepsilon})x_t + \xi_t^{\varepsilon} w_t,$$

- y_t^{ε}: contaminated observations
- x_t: core process which follows Eq. (1)
- w_t: contaminating process, independent of x_t
- ξ_t^{ε}: stationary random process

$$\xi_t^{\varepsilon} = \begin{cases} 1 & \text{with probability } \varepsilon \\ 0 & \text{with probability } (1-\varepsilon). \end{cases}$$

Outlier Models

Innovation Outliers

Innovation Outliers (IO)

a_t in Eq. (1) is replaced by

$$a_t^{\varepsilon} = a_t + \xi_t^{\varepsilon} w_t,$$

or

$$a_t^{\varepsilon} = (1 - \xi_t^{\varepsilon}) a_t + \xi_t^{\varepsilon} w_t,$$

- a_t^{ε}: contaminated innovations
- w_t: contaminating process, independent of a_t
- ξ_t^{ε}: stationary random process

$$\xi_t^{\varepsilon} = \begin{cases} 1 & \text{with probability } \varepsilon \\ 0 & \text{with probability } (1-\varepsilon). \end{cases}$$

Outlier Models

Patchy and Isolated Outliers

Outliers may also differ depending on temporal structure:

Isolated Outliers

- ξ_t^ε takes the value 1, such that at least one non-outlying observation is between two outliers (e.g. ξ_t^ε follows an independent Bernoulli distribution)

Patchy Outliers

- ξ_t^ε, $i = 1, \ldots, n$ takes the value 1 for $n_{\text{patch}} \leq n/2$ subsequent samples

Further Outlier Models

- level shifts
- change of variance
- ...

Propagation of Outliers

Propagation of Outliers

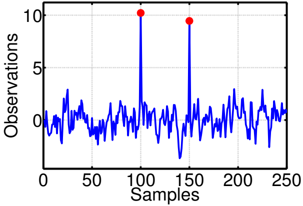
Example: Outliers in an AR(1) process

TECHNISCHE UNIVERSITÄT DARMSTADT

Example: AR(1) Process:

- $x_t = 0.5x_{t-1} + a_t, \quad t = 1, \dots, 250$

  a_t are zero mean i.i.d. Gaussian random variables with $\sigma_a = 1$
- observations given by $y_{100} = x_{100} + 10$, $y_{150} = x_{150} + 10$, $y_t = x_t$, otherwise



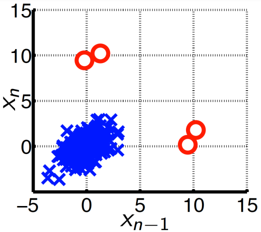
$\rightarrow$ two additive outliers

Propagation of Outliers

Example: Outliers in an AR(1) process

TECHNISCHE UNIVERSITÄT DARMSTADT

Representation of the AR(1) With Two AOs as Regression:



→ four outliers in regression representation: two vertical outliers and two 'bad' leverage points

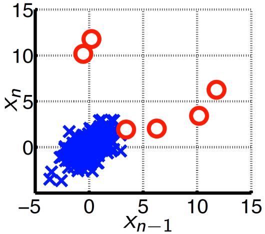
→ for $p > 1$ or ARMA, even highly robust i.i.d. regression estimators break down (reasons shown later)

Propagation of Outliers

Example: Outliers in an AR(1) process

Now Consider Two Innovations Outliers:

$a_{100} = a_{100} + 10$, $a_{150} = a_{150} + 10$ and the corresponding regression



→ two vertical outliers, but several 'good' leverage points, which lie on the tangent given by $x_n = 0.5x_{n-1}$.

→ even classical estimators give good results, since the 'good leverage points' compensate for the vertical outliers

Propagation of Outliers

Reconstructing the Innovation Series

ARMA Parameter Estimation

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad f(\mathbf{a}_n(\boldsymbol{\beta}))$$

- $\mathbf{a}_n(\boldsymbol{\beta}) = (a_1^e(\boldsymbol{\beta}), a_2^e(\boldsymbol{\beta}), \ldots, a_n^e(\boldsymbol{\beta}))$
- $a_t^e(\boldsymbol{\beta}) = \theta^{-1}(B)\phi(B)(y_t - \mu)$

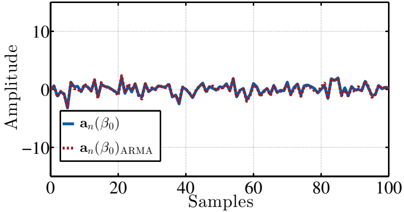
  $= y_t - \mu - \sum_{i=1}^{p} \phi_i(y_{t-i} - \mu) + \sum_{i=1}^{q} \theta_i a_{t-i}^e(\boldsymbol{\beta})$

Propagation of Outliers

ARMA(2,1) with Additive Outliers

Example: ARMA(2,1) - Reconstructing the Innovations $\mathbf{a}_n(\beta)$

- $\phi = (-0.39, -0.3)$, $\theta = 0.9$, $\mu = 0$, $\sigma = 1$
- observations follow ARMA



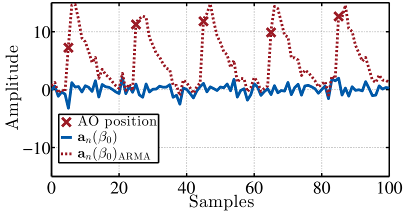
ARMA innovations estimates

Propagation of Outliers

ARMA(2,1) with Additive Outliers

Example: ARMA(2,1) - Reconstructing the Innovations $\mathbf{a}_n(\beta)$

- $\phi = (-0.39, -0.3)$, $\boldsymbol{\theta} = 0.9$, $\mu = 0$, $\sigma = 1$
- observations follow ARMA with additive outliers



ARMA innovations estimates

$\rightarrow$ propagation of outliers onto multiple innovations estimates must be prevented!

Propagation of Outliers

How to prevent it?

Two main approaches

1. filtered robust estimators
   - known since the 1970s
   - provide good results when all parameters are chosen correctly
   - implementation not straight forward
   - not tractable in terms of statistical analysis (consistency, efficiency, influence function)
2. bounded influence propagation model
   - proposed in 2009
   - auxiliary model that includes ARMA model as a special case
   - easy to implement
   - tractable in terms of statistical analysis (consistency, efficiency, influence function)

Bounded Influence Propagation Model

TECHNISCHE UNIVERSITÄT DARMSTADT

Bounded Innovation Propagation Autoregressive Moving Average (BIP-ARMA) Model, [8]

$$y_t = a_t + \mu + \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu) - \sum_{i=1}^{r} \left(\phi_i a_{t-i} + (\theta_i - \phi_i) \sigma \eta \left(\frac{a_{t-i}}{\sigma} \right) \right) \tag{2}$$

- auxiliary model to prevent propagation of outliers
- $\eta(x)$: odd, bounded and continuous function (e.g., Huber or Tukey)
  ARMA models: $\eta(x) = x$.
- σ: scale of a_t
- $r = \max(p, q)$, if $r > p$, $a_{p+1} = ... = a_r = 0$, while if $r > q$, $b_{q+1} = ... = b_r = 0$.

Bounded Influence Propagation Model

Recursion for BIP ARMA Innovations

$$a_t^b(\boldsymbol{\beta}, \sigma) = y_t - \mu - \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu) + \sum_{i=1}^{r} \left(\phi_i a_{t-i}^b(\boldsymbol{\beta}, \sigma) + (\theta_i - \phi_i)\sigma\eta\left(\frac{a_{t-i}^b(\boldsymbol{\beta}, \sigma)}{\sigma}\right)\right)$$

- recursion depends on innovations scale σ
- $\rightarrow$ How to determine σ?

Bounded Influence Propagation Model

Recursion for BIP ARMA Innovations

Eq. (1) can be written as MA(∞)

$$y_t = \mu - a_t + \sum_{i=1}^{\infty} \lambda_i \sigma \eta \left(\frac{a_{t-i}}{\sigma} \right),$$

- $\lambda_i(\boldsymbol{\beta})$: coefficients of $\phi^{-1}(B)\theta(B)$.

Now

$$\sigma^2(\boldsymbol{\beta}) = \frac{\sigma_y^2}{1 + \kappa^2 \sum_{i=1}^{\infty} \lambda_i^2(\boldsymbol{\beta})},$$

- σ_y: standard deviation of y_t
- $\kappa^2 = \text{Var}\left[\eta\left(\frac{a_t}{\sigma}\right)\right] = \text{E}\left[\left(\eta\left(\frac{a_t}{\sigma}\right) - \text{E}\left[\eta\left(\frac{a_t}{\sigma}\right)\right]\right)^2\right]$

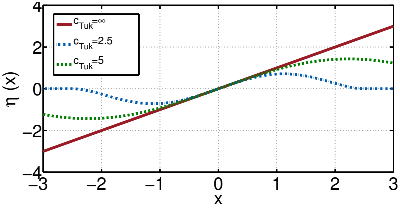
Bounded Influence Propagation Model

Example of $\eta(x)$

- Tukey's Biweight

$$\eta(x) = \begin{cases} x - 2\frac{x^3}{c_{\text{Tuk}}^2} + \frac{x^5}{c_{\text{Tuk}}^4} & |x| \leq c_{\text{Tuk}} \\ 0 & |x| > c_{\text{Tuk}}. \end{cases}$$

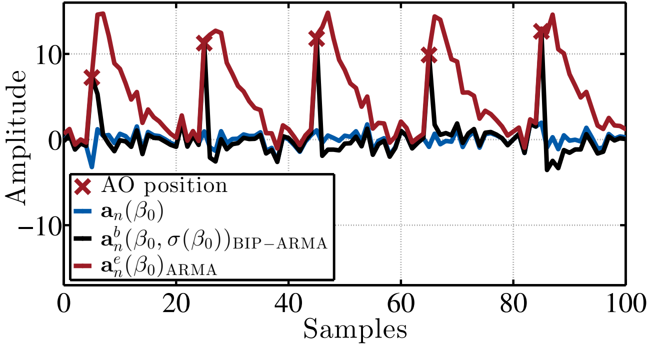
- $c_{\text{Tuk}} = \infty \rightarrow$ ARMA model



Bounded Influence Propagation Model

TECHNISCHE UNIVERSITÄT DARMSTADT

Effect of $\eta(x)$



State-Space Representation of AR(p) Process

- state equation:

$$\mathbf{x}_t = \mathbf{\Phi} \mathbf{x}_{t-1} + \mathbf{a}_t$$

  - non–observable p-dimensional state vector: $\mathbf{x}_t = [x_t, x_{t-1}, \cdots, x_{t-p+1}]^T$
  - innovations: $\mathbf{a}_t = [a_t, 0, \cdots, 0]^T$
  - state transition matrix: $\mathbf{\Phi} = \begin{bmatrix} \phi_1 & \cdots & \phi_{p-1} & \phi_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}$

- measurement equation:

$$y_t = x_t + w_t$$

  - x_t and w_t are independently distributed

Approximate Conditional Mean (ACM) Filter, [51, 22]
computes robustly filtered estimate

$$\hat{\mathbf{x}}_{t|t} = \mathbf{\Phi}\hat{\mathbf{x}}_{t-1|t-1} + \frac{\hat{\mathbf{\Sigma}}_{1,t}}{\hat{\sigma}_t}\,\psi\left(\frac{y_t - \hat{y}_{t|t-1}}{\hat{\sigma}_t}\right)$$

which is an approximation of $\mathrm{E}[\mathbf{x}_t | y_1, y_2, \ldots, y_t]$

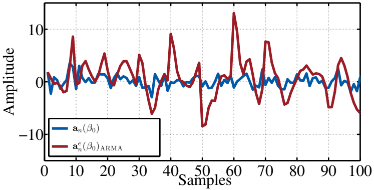
- $\psi(\cdot)$ is an odd, bounded and continuous score function [7]
- $\hat{\mathbf{\Sigma}}_{1,t}$ is the first column of $\hat{\mathbf{\Sigma}}_t$ (prediction error covariance matrix), which is computed recursively. $\hat{\sigma}_t^2$ is the first element of $\hat{\mathbf{\Sigma}}_{1,t}$
- $\hat{y}_{t|t-1}$ is the robust one step ahead predictor of y_t based on $\{y_1, \ldots, y_{t-1}\}$

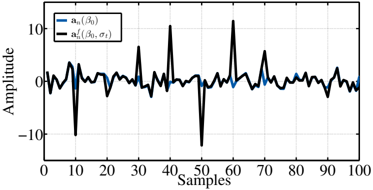
$$\hat{y}_{t|t-1} = (\mathbf{\Phi}\hat{\mathbf{x}}_{t-1|t-1})_1$$

- For a detailed description of the algorithm, see e.g. [18]

Example: Filtered Residuals for AR(2)

- $\phi^{\mathsf{T}} = (\phi_1 \; \phi_2) = (0.8 \; 0.3)$; every 10th sample AO from $\mathcal{N}(0, (10\sigma)^2)$





blue: innovations process; red: ARMA innovations estimate; black: robustly filtered residuals

- filtered residual at time t: $\quad a_t^f(\boldsymbol{\beta}, \sigma_t) = y_t - \phi^{\mathsf{T}} \hat{\mathbf{x}}_{t-1|t-1}$
- $a_t^e(\boldsymbol{\beta}_0) = y_t - \phi^{\mathsf{T}} (y_{t-1} \; y_{t-2})$

$\rightarrow$ use filtered residuals for robust parameter estimation

Robust ARMA Parameter Estimation

Robust ARMA Parameter Estimation

Overview:

- revision of some popular robust ARMA estimators
- algorithms to compute the estimates
- real data applications
- some robustness theory

Cleaned Maximum Likelihood Estimators

TECHNISCHE UNIVERSITÄT DARMSTADT

3σ cleaned ML-estimator (ML 3σ)

Simple diagnostic robust method that is frequently used among engineering practitioners.

1. *3σ rejection*, i.e. observations beyond three standard deviations are flagged as outliers.
2. ML estimation with missing data

- Justified since for $x_t \sim \mathcal{N}(\mu, \sigma^2)$, $\Pr(|x_t - \mu| < 3\sigma) = 99,73$.
- Robust estimates of the mean μ and the standard deviation σ should be used to avoid the masking effect

Median-of-Ratios-Estimator (MRE)

TECHNISCHE UNIVERSITÄT DARMSTADT

Median-of-Ratios-Estimator (MRE), [10, 12]

An ARMA(p,q) model is estimated by the MRE as follows:

1. *Fit a high order AR(p_0) using the median of y_t/y_{t-i}, where $t = i+1, i+2, \dots, n$ for $i = 1, 2, \dots, p_0$, to estimate the autocorrelation at lag i. The order $p_0 > p$ is obtained by a robust order selection criterion.*
2. *Discard the outliers by filtering the signal using a robust filter-cleaner with the estimated parameters of the high order AR(p_0) and apply a classical estimation method of ARMA models that handles missing data.*

The method offers good performance in practice and is easy to implement. However, its breakdown point is limited to 0.25.

M-estimator

TECHNISCHE UNIVERSITÄT DARMSTADT

An M-estimate is Obtained by Solving

$$\hat{\boldsymbol{\beta}}_M = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad \sum_{t=p+1}^{n} \rho\left(\frac{a_t^e(\boldsymbol{\beta})}{\hat{\sigma}_n^M(\mathbf{a}_n(\boldsymbol{\beta}))}\right) \tag{1}$$

- $a_t^e(\boldsymbol{\beta}) = y_t - \mu - \sum_{i=1}^{p} \phi_i(y_{t-i} - \mu) + \sum_{i=1}^{q} \theta_i a_{t-i}^e(\boldsymbol{\beta})$
- $\rho(x)$ is a real function with: $\rho(0) = 0$, $\rho(x) = \rho(-x)$, and $\rho(x)$ is continuous, non-constant and non-decreasing in $|x|$.
- $\hat{\sigma}_n^M(\mathbf{a}_n(\boldsymbol{\beta}))$ is an M-estimate of the innovations scale

$$\frac{1}{n-p} \sum_{t=p+1}^{n} \rho\left(\frac{a_t(\boldsymbol{\beta})}{\hat{\sigma}_n^M(\mathbf{a}_n(\boldsymbol{\beta}))}\right) = b. \tag{2}$$

- $\sup \rho(x) > b$

M-estimator

TECHNISCHE UNIVERSITÄT DARMSTADT

equivalently, with $\psi(x) = \frac{d\rho(x)}{dx}$, $\hat{\boldsymbol{\beta}}_M$ is found by solving

$$\sum_{t=p+1}^{n} \mathbf{y}_{t-1}\psi\left(\frac{a_t^e(\boldsymbol{\beta})}{\hat{\sigma}_n^M(\mathbf{a}_n(\boldsymbol{\beta}))}\right) = \mathbf{0}$$

- $\mathbf{y}_{t-1}^\top = (1, y_t, y_{t-1}, \dots, y_{t-p+1})$
- $\psi(x)$ is bounded and continuous

S-estimator

S-estimator

$$\hat{\boldsymbol{\beta}}_S = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad \hat{\sigma}(a_n^e(\boldsymbol{\beta})) \tag{1}$$

- S-estimators provide the value of $\boldsymbol{\beta}$ that minimizes an M-scale estimate as defined on the previous slide.
- finding $\hat{\boldsymbol{\beta}}_S$ requires solving a nonconvex problem for which the complexity increases with p, q
- iterative algorithms require a robust starting point

MM-estimator

TECHNISCHE UNIVERSITÄT DARMSTADT

An MM-estimate is Obtained by Solving

$$\hat{\boldsymbol{\beta}}_{MM} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{t=p+1}^{n} \rho_2 \left(\frac{a_t^e(\boldsymbol{\beta})}{\hat{\sigma}(\mathbf{a}_n^e(\hat{\boldsymbol{\beta}}_S))} \right) \tag{2}$$

- $a_t^e(\boldsymbol{\beta}) = y_t - \mu - \sum_{i=1}^{p} \phi_i(y_{t-i} - \mu) + \sum_{i=1}^{q} \theta_i a_{t-i}^e(\boldsymbol{\beta})$
- $\hat{\sigma}(\mathbf{a}_n^e(\hat{\boldsymbol{\beta}}_S))$: robust M-scale estimate of $\mathbf{a}_n^e(\hat{\boldsymbol{\beta}}_S) = (a_1^e(\hat{\boldsymbol{\beta}}_S), a_2^e(\hat{\boldsymbol{\beta}}_S), \ldots, a_n^e(\hat{\boldsymbol{\beta}}_S))$

$\rightarrow$ MM estimator is a two-step estimator that requires computing an S-estimator based on ρ_1 followed by an M-estimator with a different ρ function

τ-estimator

τ-estimator

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \quad \hat{\sigma}_n^\tau(\mathbf{a}_n(\boldsymbol{\beta})) \tag{3}$$

- τ-estimators provide the value of $\boldsymbol{\beta}$ that minimizes a robust and efficient τ-scale estimate

$$\hat{\sigma}_n^\tau(\mathbf{a}_n(\boldsymbol{\beta})) = \hat{\sigma}_n^M(\mathbf{a}_n(\boldsymbol{\beta}))\sqrt{\frac{1}{n-p}\sum_{t=p+1}^{n} \rho_2\left(\frac{a_i(\boldsymbol{\beta})}{\hat{\sigma}_n^M(\mathbf{a}_n(\boldsymbol{\beta}))}\right)}$$

- finding $\hat{\sigma}_n^\tau(\mathbf{a}_n(\boldsymbol{\beta}))$ requires solving a nonconvex problem for which the complexity increases with p, q
- iterative algorithms require a robust starting point

τ-estimator

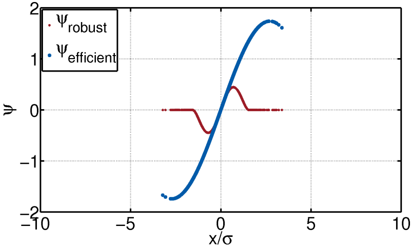
- M-estimate of the innovations scale

$$\frac{1}{n-p}\sum_{t=p+1}^{n}\rho_1\left(\frac{a_t(\beta)}{\hat{\sigma}_n^M(\mathbf{a}_n(\beta))}\right)=b. \tag{3}$$

- $\mathbf{a}_n(\beta) = (a_{p+1}(\beta), \dots, a_n(\beta))$
- $\rho_1(x)$ is a real function with: $\rho_1(0) = 0$, $\rho_1(x) = \rho_1(-x)$, and $\rho_1(x)$ is continuous, non-constant and non-decreasing in $|x|$.
- $\psi_1(x) = \frac{d\rho_1(x)}{dx}$ is bounded and continuous.
- $\sup\rho_1(x) > b$

τ-estimator

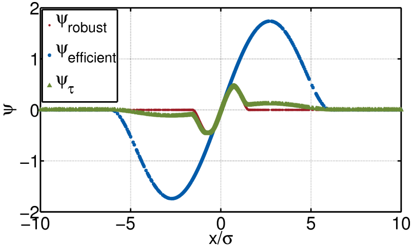
- asymptotically equivalent to a weighted sum of two M-estimates with data dependent weight



ψ-functions

τ-estimator

- asymptotically equivalent to a weighted sum of two M-estimates with data dependent weight



ψ-functions for $0 \leq \varepsilon \leq 0.5$

Robustifying Popular Robust Estimators

TECHNISCHE UNIVERSITÄT DARMSTADT

Due to Propagation of Outliers, None of These Estimators Are Robust:

- M-estimator
- S-estimator
- MM-estimator
- τ-estimator

However, they become robust as soon as the innovations $a_t^e(\boldsymbol{\beta}, \sigma)$ based on which the estimation is performed are replaced by

- $a_t^b(\boldsymbol{\beta}, \sigma)$ from the BIP-model, or
- $a_t^f(\boldsymbol{\beta}, \sigma_t)$ obtained from robust filters.

Example

Bounded Innovation Propagation τ-Estimator

Definition of the τ-Estimator Under the BIP-ARMA Model, [57]

- τ-estimate of $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$ under the BIP-ARMA model

$$\hat{\boldsymbol{\beta}}_{\tau}^{b} = \arg\min_{\boldsymbol{\beta}\in\mathcal{B}} \hat{\sigma}_n^{\tau}(\mathbf{a}_n^b(\boldsymbol{\beta}, \hat{\sigma}(\boldsymbol{\beta}))), \tag{3}$$

- τ-estimate of the scale of $\mathbf{a}_n^b(\boldsymbol{\beta}, \hat{\sigma}(\boldsymbol{\beta}))$, which can be computed recursively from (2)

Example

Bounded Innovation Propagation τ-Estimator

Final BIP τ-estimator, [57]

$$\hat{\boldsymbol{\beta}}_\tau^* = \begin{cases} \hat{\boldsymbol{\beta}}_\tau & \text{if} \quad \hat{\sigma}_n^\tau(\mathbf{a}_n(\hat{\boldsymbol{\beta}}_\tau)) < \hat{\sigma}_n^\tau(\mathbf{a}_n^b(\hat{\boldsymbol{\beta}}_\tau^b, \hat{\sigma}(\hat{\boldsymbol{\beta}}_\tau^b))) \\ \hat{\boldsymbol{\beta}}_\tau^b & \text{else.} \end{cases} \tag{4}$$

- in [57], it is shown that for (outlier-free) ARMA models, it asymptotically holds that $\hat{\sigma}_n^\tau(\mathbf{a}_n(\hat{\boldsymbol{\beta}}_\tau)) < \hat{\sigma}_n^\tau(\mathbf{a}_n^b(\hat{\boldsymbol{\beta}}_\tau^b, \hat{\sigma}(\hat{\boldsymbol{\beta}}_\tau^b)))$.

$\rightarrow$ when the data follows an (outlier-free) ARMA, the ARMA-based estimate is used.

Algorithm for AR(p)

Bounded Influence Propagation τ

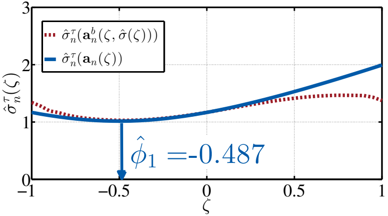
To compute $\hat{\beta}^*_\tau$ for the AR(p) model, [57] proposes a robust Durbin-Levinson type algorithm, where the parameters are recursively found for $m = 1, \dots, p$.

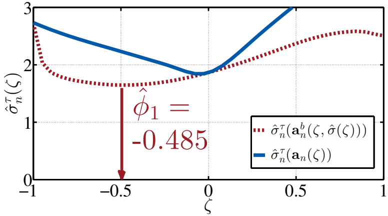
For AR(1), proceed as follows

- define a grid $\zeta^0 = -0.99 : \Delta_{\zeta^0} : 0.99$
- compute AR(1) innovations from ARMA and BIP-ARMA model, i.e. $\mathbf{a}_n(\zeta^0), \mathbf{a}_n^b(\zeta^0, \hat{\sigma}(\zeta^0))$
- compute corresponding τ-scale estimates $\hat{\sigma}_\tau(\mathbf{a}_n(\zeta^0))$, and $\hat{\sigma}_\tau(\mathbf{a}_n^b(\zeta^0, \hat{\sigma}(\zeta^0)))$
- Estimate for AR(1) is given by $\hat{\phi}_1 = \underset{\zeta}{\operatorname{argmin}} \left\{\hat{\sigma}_\tau(\mathbf{a}_n(\zeta)), \hat{\sigma}_\tau(\mathbf{a}_n^b(\zeta, \hat{\sigma}(\zeta)))\right\}$.

Algorithm for AR(p)

Bounded Influence Propagation τ





Example of finding $-1 < \zeta < 1$ which minimizes $\hat{\sigma}_n^\tau(\mathbf{a}_n(\zeta))$and $\hat{\sigma}_n^\tau(\mathbf{a}_n^b(\zeta, \hat{\sigma}(\zeta)))$ for an AR(1) process with $\phi_1 = -0.5$ and $\sigma = 1$. (left) $y_t = x_t$ clean data example; (right) 10 % equally spaced AOs of amplitude 10.

Algorithm for AR(p)

Bounded Influence Propagation τ

For a general AR(p) process, the parameters are found recursively for $m = 2, \dots, p$ by minimizing

$$\hat{\phi}_{m,m} = \underset{\zeta}{\text{argmin}} \left\{ \hat{\sigma}_\tau(\mathbf{a}_n(\zeta)), \hat{\sigma}_\tau(\mathbf{a}_n^b(\zeta, \hat{\sigma}(\zeta))) \right\} \tag{5}$$

at each order m in the same manner described for the AR(1), with the help of the Durbin-Levinson recursion:

$$\hat{\phi}_{m,m} = \begin{cases} \zeta & \text{if} \quad i = m \\ \hat{\phi}_{m-1,i} - \zeta \hat{\phi}_{m-1,m-i} & \text{if} \quad 1 \leq i \leq m-1 \end{cases} \tag{6}$$

Algorithm for ARMA(p, q)

Bounded Influence Propagation τ

TECHNISCHE UNIVERSITÄT DARMSTADT

ARMA parameter estimation is more challenging than AR

- nonconvex problems, i.e. minimizing τ-scale under the ARMA and BIP ARMA must be solved
- most important question: how to robustly find a starting point
- grid search not feasible for larger values of p, q

One possibility presented in [57] is described in the sequel.

First, starting from the BIP model, the robust one step prediction of y_t can be computed recursively for $t \geq p + 1$ via

$$\hat{y}_t = \mu + \sum_{i=1}^{p} \phi_i \left(y_{t-i} - \mu a_t^b(\hat{\boldsymbol{\beta}}, \hat{\sigma}) + \hat{\sigma}\eta \left(\frac{a_{t-i}^b(\hat{\boldsymbol{\beta}}, \hat{\sigma})}{\hat{\sigma}} \right) \right) - \sum_{i=1}^{q} \theta_i \hat{\sigma}\eta \left(\frac{a_{t-i}^b(\hat{\boldsymbol{\beta}}, \hat{\sigma})}{\hat{\sigma}} \right). \quad (7)$$

Real Data Example

Bounded Influence Propagation τ Artifact Cleaning

TECHNISCHE UNIVERSITÄT DARMSTADT

With the BIP model, we obtain outlier-cleaned observations for $t \geq p + 1$ by computing

$$y_t^* = y_t - a_t^b(\hat{\beta}, \hat{\sigma}) + \hat{\sigma}\eta\left(\frac{a_t^b(\hat{\beta}, \hat{\sigma})}{\hat{\sigma}}\right). \tag{8}$$

Robust Starting Point Procedure

- apply an AR(p) approximation to (8)
- p must be chosen sufficiently large to well approximate the ARMA
- compute estimates via Durbin-Levinson algorithm.

This produces outlier-cleaned observations, for which we can use any classical ARMA parameter estimator, e.g. [50] to obtain a robust starting point for the nonlinear LS algorithm that minimizes the τ-scales under the ARMA and BIP ARMA models.

Meausures of Robustness

Influence Function

TECHNISCHE UNIVERSITÄT DARMSTADT

Dependent Data Influence Function [47]
Directional derivative at $F(x)$, i.e.

$$\mathrm{IF}(\{F(x, \xi^{\varepsilon}, w)\}; \hat{\beta}_{\infty}) = \lim_{\downarrow \varepsilon}(\hat{\beta}_{\infty}(F(y^{\varepsilon})) - \hat{\beta}_{\infty}(F(x))) = \frac{\partial}{\partial \varepsilon}\hat{\beta}_{\infty}(F(y^{\varepsilon}))\big|_{\varepsilon=0},$$

provided that the limit exists.

- $F(x)$, $F(w)$, $F(\xi^{\varepsilon})$ and $F(y^{\varepsilon})$ are the cdfs of x_t, w_t, ξ^{ε} and y_t^{ε}, respectively.
- $F(x, \xi^{\varepsilon}, w)$ is the joint distribution of x_t, w_t, ξ^{ε}.

→ for dependent data, IF changes depending on the outlier model (i.i.d.: contamination process represented by a Dirac distribution)

- two definitions exist that are mathematically related

Influence Function

TECHNISCHE UNIVERSITÄT DARMSTADT

Dependent Data Influence Function [47]

Defined for functionals which may be computed as a solution to the estimating equation

$$\int \tilde{\psi}(\mathbf{y}_t, \hat{\beta}) dF(\mathbf{y}_t) = 0.$$

- $\mathbf{y}_t$: observations
- $F(\mathbf{y}_t)$: distribution of the observations
- $\tilde{\psi}$: function of the observations and the estimator
- class is quite large and contains both classical and robust parameter estimators, e.g. the M-estimators, the generalized M-estimators and estimators based on residual autocovariances (RA-estimators)
- However, computation of IF has only been performed for AR(1) and MA(1) models.

Influence Function

Example

TECHNISCHE UNIVERSITÄT DARMSTADT

IF of τ-Estimator for AR(1) With AO [57]

- Let y_t^ε follow the AO model with x_t satisfying and AR with $p = 1$ and $\mu = 0$.
- Further, let the ξ_t^ε be an independently distributed 0-1 sequence that is independent of x_t and w_t.

Then, under the assumptions given in [57]

$$\mathrm{IF}(F(w), \hat{\boldsymbol{\beta}}_\tau, \phi) = \frac{(1-\phi_1^2)^{1/2}}{\mathcal{E}_0} \mathrm{E}\left[(x_0 + w_0)(1-\phi_1^2)^{1/2} \psi_\tau\left(a_1 - \phi_1 w_0\right)\right] \qquad (9)$$

- $\psi_\tau(x) = W_n(\boldsymbol{\beta})\psi_1\left(\frac{a_t(\boldsymbol{\beta})}{\hat{\sigma}_n^M(\mathbf{a}_n(\boldsymbol{\beta}))}\right) + \psi_2\left(\frac{a_t(\boldsymbol{\beta})}{\hat{\sigma}_n^M(\mathbf{a}_n(\boldsymbol{\beta}))}\right)$, $\quad W_n(\boldsymbol{\beta})$ is derived in [57]
- $\mathcal{E}_0 = E\left[\nu^2 \left.\frac{\partial(\psi_\tau(x))}{\partial x}\right|_{x=u}\right] \neq 0$,
- ν and u: independent standard normal random variables.

Influence Function

Example

IF of τ-Estimator for AR(1) With AO [Muma 2016]

For $P(w_t = c_w) = 1$, where c_w is a constant, the IF has the appealing heuristic interpretation of displaying the influence of a contamination value c_w on the estimator, similarly to Hampel's definition [48] for iid data.

Example: IF for AR(1)

AR(1) with $\phi = -0.5$ for independent AOs of magnitude c_w

$$\rho_2(x) = \begin{cases} 0.5x^2 & \text{if } |x| \leq 2 \\ 0.002x^8 - 0.052x^6 + 0.432x^4 - 0.972x^2 + 1.792 & \text{if } 2 < |x| \leq 3 \\ 3.25 & |x| > 3, \end{cases}$$

$\rho_1(x) = \rho_2(x/c_1)$, with $c_1 = 0.4050$ and $\eta(x) = d\rho_2(x)/dx$.

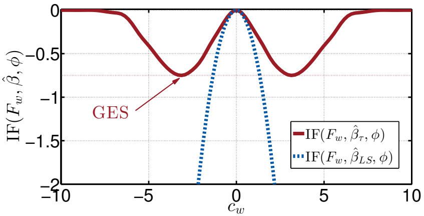
Influence Function

Example

IF of τ-Estimator for AR(1) With AO [Muma 2016]

For $P(w_t = c_w) = 1$, where c_w is a constant, the IF has the appealing heuristic interpretation of displaying the influence of a contamination value c_w on the estimator, similarly to Hampel's definition [48] for iid data.

Example: IF for AR(1)



IFs of τ-estimator and least-squares (LS) estimator. GES is the gross-error-sensitivity.

Maximum Bias Curve

Example

TECHNISCHE UNIVERSITÄT DARMSTADT

In Practice: MBC Usually Approximated Via Monte Carlo Simulations

[7, 10, 12]

$$\mathrm{MBC}(\varepsilon) = \sup_{c_w} \left| \hat{\boldsymbol{\beta}}_n(\varepsilon, c_w) - \boldsymbol{\beta} \right|$$

- The approximation is done by choosing for $\mathrm{MBC}(\varepsilon)$ the worst-case estimate of $\boldsymbol{\beta}$ over all Monte Carlo runs for a given contamination probability ε.
- c_w is a deterministic value that is varied on a grid such that for each value of c_w, the distribution of w_t (see (17)) is given by $\Pr(w_t = -c_w) = \Pr(w_t = c_w) = 0.5$.

Maximum Bias Curve

Example

More General Definition: Quantile Bias Curve (QBC)

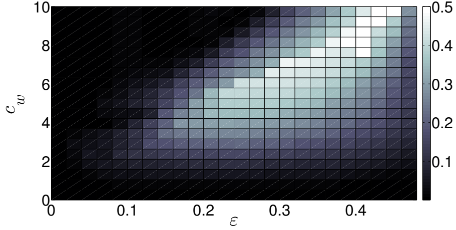
$$\mathrm{QBC}\alpha(\varepsilon) = Q_\alpha \left\{ \left| \hat{\beta}_n(\varepsilon, c_w) - \beta \right| \right\}. \tag{10}$$

- states that α percent of the sorted data is to the left of Q_α.
- E.g. QBC75(ε) represents the MBC obtained in 75 % of the Monte Carlo runs for varying c_w and fixed ε. QBC50(ε) corresponds to the Median BC(ε) and QBC100(ε) is the MBC(ε).

Maximum Bias Curve

Example: MBC analysis for BIP τ- estimator [57]

Maximum Bias For a Given Pair of (c_w, ε) for BIP τ- estimator



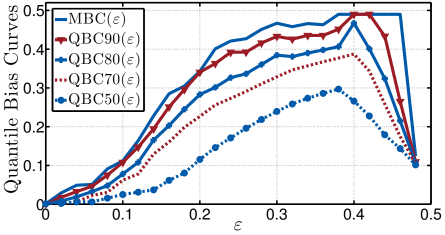
- $\phi = 0.5$
- asymptotic value was approximated using $n = 10\,000$

Maximum Bias Curve

Example: MBC analysis for BIP τ- estimator [57]

TECHNISCHE UNIVERSITÄT DARMSTADT

QBC Obtained Assuming Worst Possible c_w For a Fixed ε



- $\phi = 0.5$
- asymptotic value was approximated using $n = 10\,000$

Breakdown Point

TECHNISCHE UNIVERSITÄT DARMSTADT

Breakdown Point for Dependent Data

- extending the notion from iid to dependent data not straight forward
- in dependent data models, e.g. ARMA, parameter space is bounded
- effect of outliers is more complicated than for location, scale or regression models
- BP will depend on type of contamination
- bias will depend on the signal model and the contamination

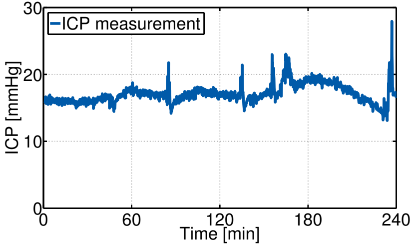
General definition by Genton and Lucas 2003 [25]

- breakdown occurs for some contamination rate ε_0, for which further increasing the value of ε does not increase the range of values taken on by the estimate over the contamination neighborhood.
- Loosely speaking: estimator is “stuck” at some value.

TECHNISCHE UNIVERSITÄT DARMSTADT

Real Data Example

Intracranial Pressure (ICP)



four hour excerpt of an ICP measurement

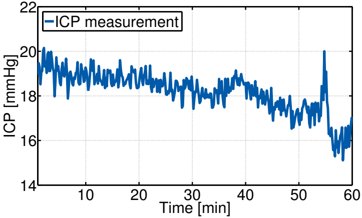
artifacts and nonstationary signal

Real Data Example

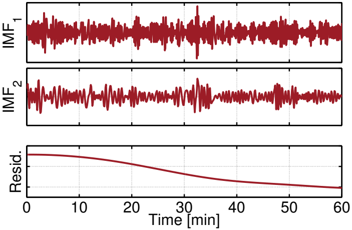
Bounded Influence Propagation τ Artifact Cleaning

TECHNISCHE UNIVERSITÄT DARMSTADT

Intracranial Pressure (ICP)



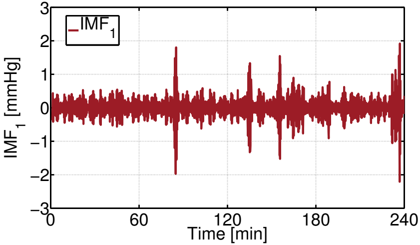
excerpt of ICP signal



intrinsic mode functions and residual

nonstationary $\rightarrow$ empirical mode decomposition $\rightarrow$ intrinsic mode functions (IMF)

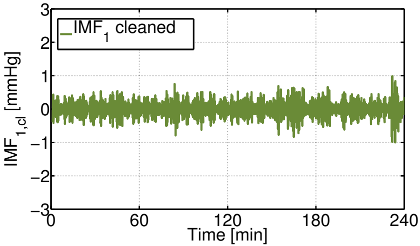
Intracranial Pressure (ICP)



first intrinsic mode function (IMF)

- ARMA BIP-τ parameter estimation and model order selection $\rightarrow$ ARMA(2,1)

Intracranial Pressure (ICP)



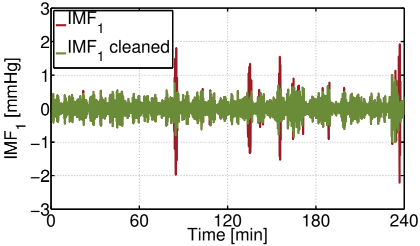
first intrinsic mode function (IMF)

- ARMA BIP-τ artifact removal $y_t^* = y_t - a_t^b(\hat{\beta}, \hat{\sigma}) + \hat{\sigma}\eta\left(\frac{a_t^b(\hat{\beta},\hat{\sigma})}{\hat{\sigma}}\right)$
  $\hat{p} = 2$, $\hat{q} = 1$, $\hat{\phi} = (0.1533, 0.0671)^T$, $\theta = -0.5283$

Real Data Example

Bounded Influence Propagation τ Artifact Cleaning

Intracranial Pressure (ICP)

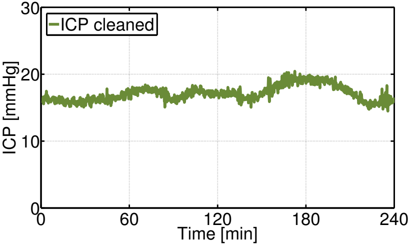


first intrinsic mode function (IMF)

- ARMA BIP-τ artifact removal $y_t^* = y_t - a_t^b(\hat{\boldsymbol{\beta}}, \hat{\sigma}) + \hat{\sigma}\eta\left(\frac{a_{t-i}^b(\hat{\boldsymbol{\beta}},\hat{\sigma})}{\hat{\sigma}}\right)$

  $\hat{p} = 2$, $\hat{q} = 1$, $\hat{\boldsymbol{\phi}} = (0.1533, 0.0671)^T$, $\theta = -0.5283$

Intracranial Pressure (ICP)



four hour excerpt of an ICP measurement

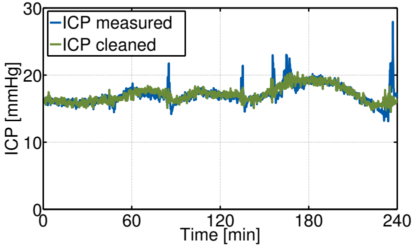
- ARMA BIP-τ artifact removal for all IMFs $\rightarrow$ back transform

Real Data Example

Bounded Influence Propagation τ Artifact Cleaning

TECHNISCHE UNIVERSITÄT DARMSTADT

Intracranial Pressure (ICP)



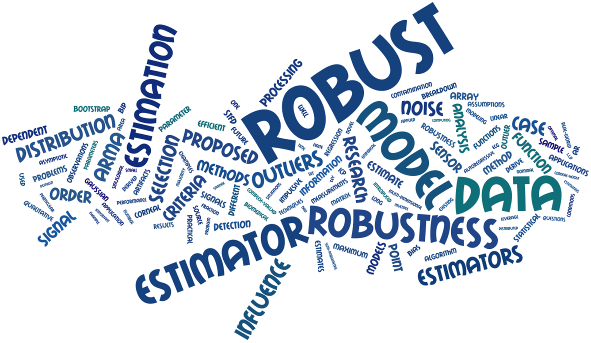
four hour excerpt of an ICP measurement

- ARMA BIP-τ artifact removal for all IMFs $\rightarrow$ back transform

M. Muma, A.M. Zoubir, Robust ARMA Parameter Estimation by Bounded Innovation τ–estimation., submitted to IEEE Trans. Signal Proc. 2016.
M. Muma, Robust Model Selection for ARMA models based on the bounded innovation τ–estimator., IEEE SSP'14, pp. 428–431, 2014.

Open Areas and Future Trends

Isaac Newton “What we know is a drop, what we don’t know is an ocean.”



Open Areas and Future Trends

- defining robustness for dependent data still not complete
- measuring robustness in higher order models
- sparsity and dependent data
- computing estimates in reasonable time
- multivariate dependent data
- characterizing (directional) coherence robustly

Thanks for your attention!

[1] R.S. Tsay,
"Outliers, level shifts, and variance changes in time series,"
J. Forecasting, vol. 7, no. 1, pp. 1–20, Jan 1988.

[2] S.J. Deutsch, J. E. Richards, and J.J. Swain,
"Effects of a single outlier on ARMA identification,"
Commun. Stat. Theory, vol. 19, no. 6, pp. 2207–2227, 1990.

[3] G.M. Ljung,
"On outlier detection in time series,"
J. Roy. Stat. Soc. B, pp. 559–567, 1993.

[4] C. Chen and L.-M. Liu,
"Joint estimation of model parameters and outlier effects in time series,"
J. Am. Stat. Assoc., vol. 88, no. 421, pp. 284–297, 1993.

[5] D.W. Shin, S. Sarkar, and J.H. Lee,
"Unit root tests for time series with outliers,"
Stat. Probabil. Lett., vol. 30, no. 3, pp. 189–197, 1996.

[6] X. de Luna and M. G. Genton,
"Robust simulation-based estimation of ARMA models,"
J. Comput. Graph. Stat., vol. 10, no. 2, pp. 370–387, 2001.

[7] R. A. Maronna, R. D. Martin, and V. J. Yohai,
Robust Statistics, Theory and Methods,
John Wiley & Sons, Ltd, 2006.

[8] N. Muler, D. Peña, and V. J. Yohai,
"Robust estimation for ARMA models,"
Ann. Statist., vol. 37, no. 2, pp. 816–840, 2009.

[9] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma,
"Robust estimation in signal processing: a tutorial-style treatment of fundamental concepts,"
IEEE Signal Process. Mag., vol. 29, no. 4, pp. 61–80, Jul 2012.

[10] Y. Chakhchoukh, P. Panciatici, and P. Bondon,
"Robust estimation of SARIMA models: Application to short-term load forecasting,"
in *In Proc. IEEE Workshop Statist. Signal Proces. (SSP 2009), Cardiff, UK*, Aug 2009.

[11] Y. Chakhchoukh,
Contribution to the estimation of SARIMA (application to short-term forecasting of electricity consumption),
Ph.D. thesis, Université de Paris-Sud, Faculté des Sciences d'Orsay, Essonne, 2010.

[12] Y. Chakhchoukh, P. Panciatici, and L. Mili,
"Electric load forecasting based on statistical robust methods,"
IEEE Trans. Power Syst., vol. 26, no. 3, pp. 982–991, Mar. 2010.

[13] B. Andrews,
"Rank-based estimation for autoregressive moving average time series models,"
J. Time Ser. Anal., vol. 29, no. 1, pp. 51–73, 2008.

[14] H. Louni,
"Outlier detection in ARMA models,"
J. Time Series Anal., vol. 29, no. 6, pp. 1057–1065, 2008.

[15] B. Han, M. Muma, M. Feng, and A. M. Zoubir,
"An online approach for intracranial pressure forecasting based on signal decomposition and robust statistics,"
in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 6239–6243.

[16] M. Muma,
"Robust model order selection for ARMA models based on the bounded innovation propagation τ-estimator,"
in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, 2014, pp. 428–431.

[17] F. Strasser, M. Muma, and A. M. Zoubir,
"Motion artifact removal in ECG signals using multi-resolution thresholding,"
in *In Proc. European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 899–903.

[18] B. Spangl and R. Dutter,
"Estimating spectral density functions robustly,"
REVSTAT-Statst. J., vol. 5, no. 1, pp. 41–61, 2007.

[19] T. Schäck, C. Sledz, M. Muma, and A. M. Zoubir,
"A new method for heart rate monitoring during physical exercise using photoplethysmographic signals,"
in *23rd European Signal Processing Conference (EUSIPCO)*, Aug 2015, pp. 2666–2670.

[20] F. F. Molinares, V. A.Reisen, and F. Cribari-Neto,
"Robust estimation in long-memory processes under additive outliers,"
J. Stat. Plan. Infer., vol. 139, no. 8, pp. 2511–2525, 2009.

[21] Y. S. Kharin and V. A. Voloshko,
"Robust estimation of AR coefficients under simultaneously influencing outliers and missing values,"
J. Stat. Plan. Infer., vol. 141, no. 9, pp. 3276 – 3288, 2011.

[22] R. D. Martin and D. J. Thomson,
"Robust-resistant spectrum estimation,"
Proc. IEEE, vol. 70, no. 9, pp. 1097–1115, Sept 1982.

[23] S. A. Kassam and V. Poor,
"Robust techniques for signal processing: a survey,"
Proc. IEEE, vol. 73, no. 3, pp. 433–481, Mar 1985.

[24] O. H. Bustos and V. J. Yohai,
"Robust estimates for ARMA models,"
J. Am. Statist. Assoc., vol. 81, no. 393, pp. 155–168, 1986.

[25] M. G. Genton and A. Lucas
"Comprehensive definitions of breakdown points for independent and dependent observations,"
J Royal Statist. Soc. Ser. B, vol. 65, no. 1, pp. 81–94, 2003.

[26] J. G. Gonzalez and G. R. Arce,
"Optimality of the myriad filter in practical impulsive-noise environments,"
IEEE Trans. Signal Process., vol. 49, no. 2, pp. 438–441, Feb 2001.

[27] Y. Yang, H. He, and G. Xu,
"Adaptively robust filtering for kinematic geodetic positioning,"
J. Geodesy, vol. 75, no. 2, pp. 109–116, 2001.

[28] L. Mili, M. G. Cheniae, and P. J. Rousseeuw,
"Robust state estimation of electric power systems,"
IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 41, no. 5, pp. 349–358, May 2002.

[29] A. D. McQuarrie and C.-L. Tsai,
"Outlier detections in autoregressive models,"
J. Comput. Graph. Stat., vol. 12, no. 2, pp. 450–471, 2003.

[30] P. Chareka, F. Matarise, and R. Turner,
"A test for additive outliers applicable to long-memory time series,"
J. Econ. Dyn. Control, vol. 30, no. 4, pp. 595 – 621, 2006.

[31] T. C. Aysal and K. E. Barner,
"Meridian filtering for robust signal processing,"
IEEE Trans. Signal Process., vol. 55, no. 8, pp. 3349–3962, Aug 2007.

[32] K. Liang, X. Wang, and T. H. Li,
"Robust discovery of periodically expressed genes using the Laplace periodogram,"
BMC Bioinform., vol. 10, no. 1, pp. 1–15, 2009.

[33] R. Nunkesser, R. Fried, K. Schettlinger, and U. Gather,
"Online analysis of time series by the Q_n estimator,"
Comput. Stat. Data An., vol. 53, no. 6, pp. 2354–2362, 2009.

[34] H. Dong, Z. Wang, and H. Gao,
"Robust H_∞ filtering for a class of nonlinear networked systems with multiple stochastic communication delays and packet dropouts,"
IEEE Trans. Signal Process., vol. 58, no. 4, pp. 1957–1966, Apr 2010.

[35] T. H. Li,
"A nonlinear method for robust spectral analysis,"
IEEE Trans. Signal Process., vol. 58, no. 5, pp. 2466–2474, May 2010.

[36] M. A. Gandhi and L. Mili,
"Robust Kalman filter based on a generalized maximum-likelihood-type estimator,"
IEEE Trans. Signal Process., vol. 58, no. 5, pp. 2509–2520, May 2010.

[37] C. Becker, R. Fried, and S. Kuhnt,
Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather,
Springer Science & Business Media, 2014.

[38] H. Dehling, R. Fried, and M. Wendler,
"A robust method for shift detection in time series,"
arXiv preprint arXiv:1506.03345, 2015.

[39] A. Dürre, R. Fried, and T. Liboschik,
"Robust estimation of (partial) autocorrelation,"
Wiley Interdisciplinary Reviews: Computational Statistics, vol. 7, no. 3, pp. 205–222, 2015.

[40] R. C. Molinari, S. Guerrier, and M.-P. Victoria-Feser,
"Robust inference for time series models: a wavelet-based framework,"
Archive ouverte UNIGE (Preprint), 2015.

[41] P. J. Huber and E. M. Ronchetti,
Robust Statistics, vol. 2,
John Wiley & Sons, Inc., Publication, 2009.

[42] V.J. Yohai and R.H. Zamar,
"High breakdown-point estimates of regression by means of the minimization of an efficient scale,"
J. Amer. Statist. Assoc., vol. 83, no. 402, pp. 406–413, 1988.

[43] V. J. Yohai,
"High breakdown-point and high efficiency estimates for regression,"
Ann. Statist., vol. 15, pp. 642–656, 1987.

[44] N. Muler, D. Peña, and V. J. Yohai,
"Robust estimation for ARMA models,"
Tech. Rep., Universidad Torcuato di Tella, Universidad Carlos III de Madrid and Universidad de Buenos Aires and CONICET, 2007.

[45] V. Yohai and R. Zamar,
"High breakdown-point estimates of regression by means of the minimization of an efficient scale,"
Tech. Rep. 84, University of Washington,, Aug 1986.

[46] N. Muler and Yohai V. J.,
"Robust estimates for ARCH processes,"
J. Time Ser. Anal., vol. 23, no. 3, pp. 341–375, 2002.

[47] R. D. Martin and V. J. Yohai,
"Influence functionals for time series,"
Ann. Statist., vol. 14, no. 3, pp. 781–818, 1986.

[48] F. R. Hampel,
"The influence curve and its role in robust estimation,"
J. Amer. Statist. Assoc., vol. 40, no. 1, pp. 375–382, 1974.

[49] Y. Chakhchoukh,
"A new robust estimation method for ARMA models,"
IEEE Trans. Signal Process., vol. 58, no. 7, pp. 3512–3522, Jul 2010.

[50] R. H. Jones,
"Maximum likelihood fitting of ARMA models to time series with missing observations,"
Technometrics, vol. 22, no. 3, pp. 389–395, 1980.

[51] C. Masreliez,
"Approximate non-Gaussian filtering with linear state and observation relations,"
IEEE Trans. Autom. Control, vol. 20, no. 1, pp. 107–110, Feb 1975.

[52] A. D. R. McQuarrie and C.-L. Tsai,
Regression and Time Series Model Selection,
World Scientific Publishing Co. Pte. Ltd., 1998.

[53] R. Moses, P. Stoica, B. Friedlander, and T. Söderström,
"An efficient linear method of ARMA spectral estimation,"
in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 1987, pp. 2077–2080.

[54] J. Pan and W. J. Tompkins,
"A real-time QRS detection algorithm,"
IEEE Trans. Biomed. Eng., vol. 1, no. 3, pp. 230–236, Mar 1985.

[55] G. D. Clifford,
Signal processing methods for heart rate variability,
Ph.D. thesis, Department of Engineering Science, University of Oxford, 2002.

[56] A. J. Q. Sarnaglia, V. A. Reisen, and P. Bondon,
"Periodic ARMA models: Application to particulate matter concentrations,"
in *In Proc. European Signal Processing Conference (EUSIPCO)*, Aug 2015, pp. 2181–2185.

[57] M. Muma and A. M. Zoubir,
"Bounded Influence Propagation τ-Estimation: A New Robust Method for ARMA Models,"
IEEE Trans. Signal Process., submitted, 2016.