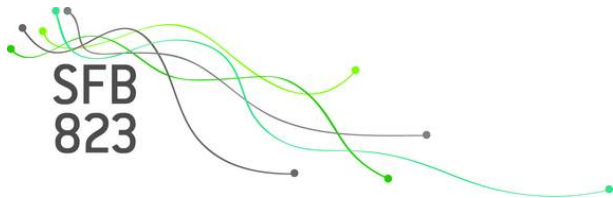


Data-driven window width adaption for robust online moving window regression

Matthias Borowski

Fakultät Statistik, TU Dortmund

COMPSTAT 2010, Paris



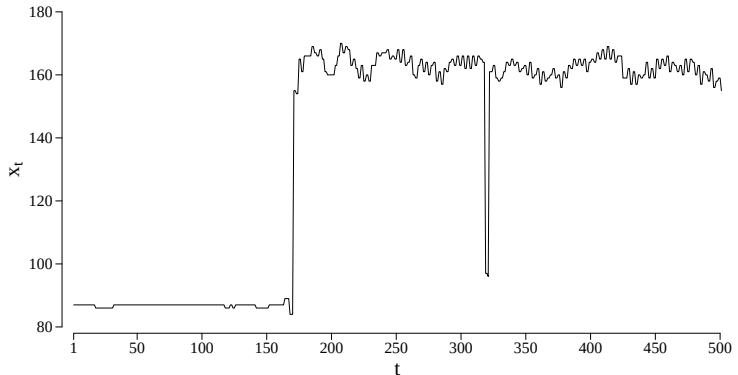


Situation

Data-driven
window
width
adaption
for robust
online
moving
window
regression

Matthias
Borowski

Online-monitoring time series: systolic blood pressure



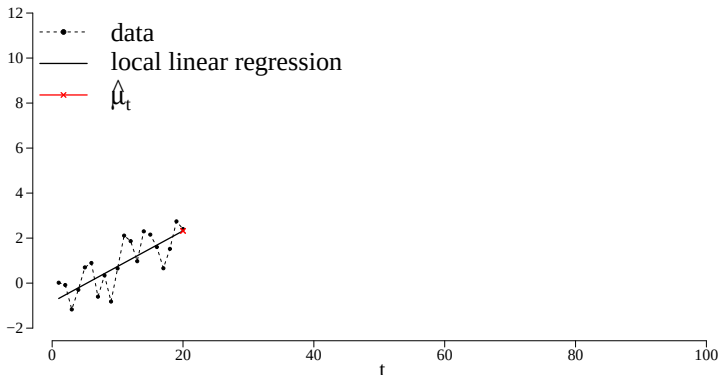
General assumption:

$$\begin{array}{ccccccc} X_t & = & \mu_t & + & \varepsilon_t & + & \eta_t \\ \text{data} & = & \text{signal} & + & \text{noise} & + & \text{outliers} \end{array}$$

Online filtering by Repeated Median (RM) regression

Data-driven
window
width
adaption
for robust
online
moving
window
regression

Matthias
Borowski



RM regression (Siegel, 1982):

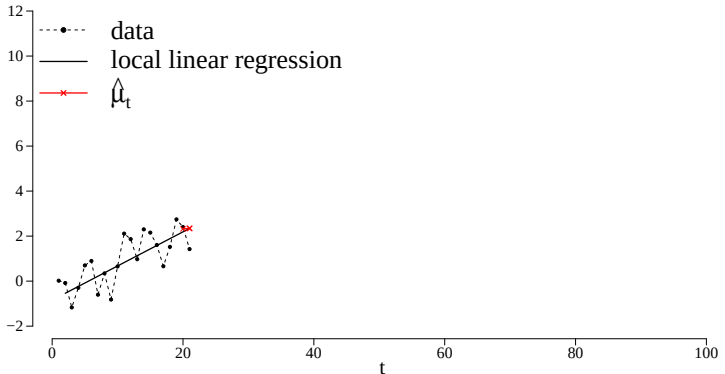
$$\text{Slope} \quad \hat{\beta}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ \operatorname{med}_{j \neq i, j \in \{1, \dots, n\}} \frac{x_{t-n+i} - x_{t-n+j}}{i - j} \right\}$$

$$\text{Level} \quad \hat{\mu}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ x_{t-n+i} - \hat{\beta}_t \cdot (i - n) \right\}$$

Online filtering by Repeated Median (RM) regression

Data-driven
window
width
adaption
for robust
online
moving
window
regression

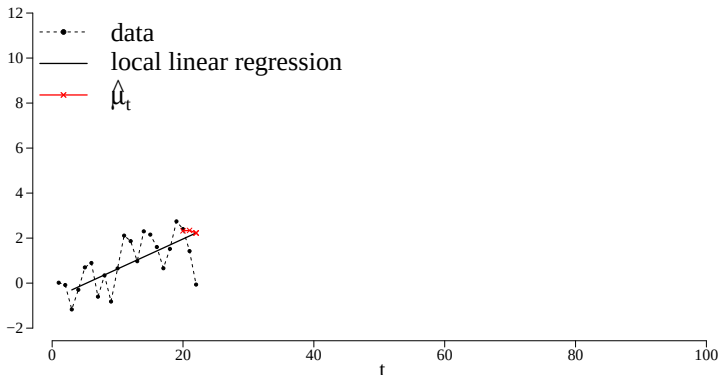
Matthias
Borowski



RM regression (Siegel, 1982):

$$\text{Slope} \quad \hat{\beta}_t = \text{med}_{i \in \{1, \dots, n\}} \left\{ \text{med}_{j \neq i, j \in \{1, \dots, n\}} \frac{x_{t-n+i} - x_{t-n+j}}{i - j} \right\}$$

$$\text{Level} \quad \hat{\mu}_t = \text{med}_{i \in \{1, \dots, n\}} \left\{ x_{t-n+i} - \hat{\beta}_t \cdot (i - n) \right\}$$



RM regression (Siegel, 1982):

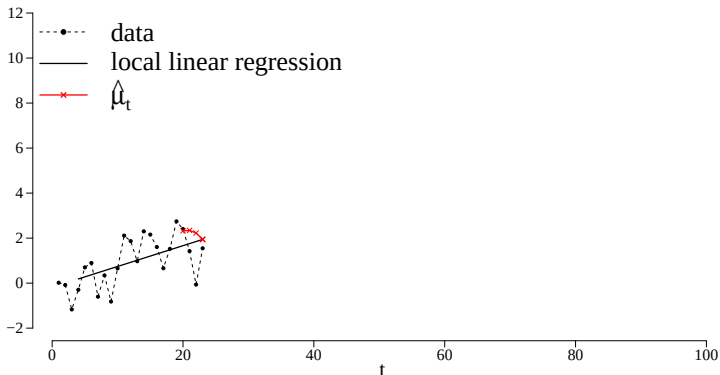
$$\text{Slope} \quad \hat{\beta}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ \operatorname{med}_{j \neq i, j \in \{1, \dots, n\}} \frac{x_{t-n+i} - x_{t-n+j}}{i - j} \right\}$$

$$\text{Level} \quad \hat{\mu}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ x_{t-n+i} - \hat{\beta}_t \cdot (i - n) \right\}$$

Online filtering by Repeated Median (RM) regression

Data-driven
window
width
adaption
for robust
online
moving
window
regression

Matthias
Borowski



RM regression (Siegel, 1982):

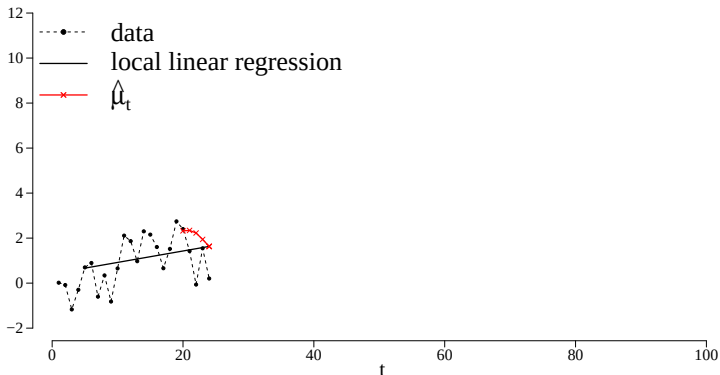
$$\text{Slope} \quad \hat{\beta}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ \operatorname{med}_{j \neq i, j \in \{1, \dots, n\}} \frac{x_{t-n+i} - x_{t-n+j}}{i - j} \right\}$$

$$\text{Level} \quad \hat{\mu}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ x_{t-n+i} - \hat{\beta}_t \cdot (i - n) \right\}$$

Online filtering by Repeated Median (RM) regression

Data-driven
window
width
adaption
for robust
online
moving
window
regression

Matthias
Borowski



RM regression (Siegel, 1982):

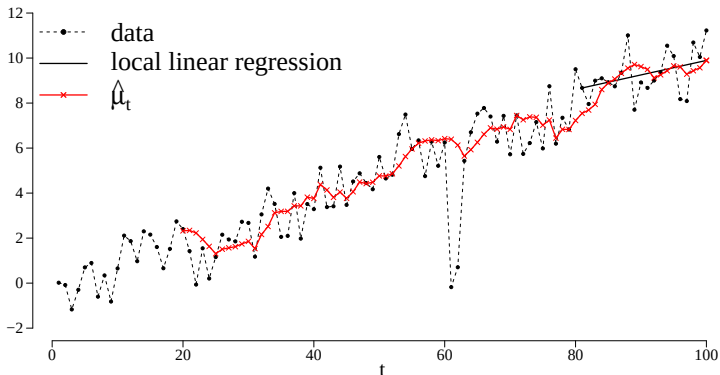
$$\text{Slope} \quad \hat{\beta}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ \operatorname{med}_{j \neq i, j \in \{1, \dots, n\}} \frac{x_{t-n+i} - x_{t-n+j}}{i - j} \right\}$$

$$\text{Level} \quad \hat{\mu}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ x_{t-n+i} - \hat{\beta}_t \cdot (i - n) \right\}$$

Online filtering by Repeated Median (RM) regression

Data-driven
window
width
adaption
for robust
online
moving
window
regression

Matthias
Borowski



RM regression (Siegel, 1982):

$$\text{Slope} \quad \hat{\beta}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ \operatorname{med}_{j \neq i, j \in \{1, \dots, n\}} \frac{x_{t-n+i} - x_{t-n+j}}{i - j} \right\}$$

$$\text{Level} \quad \hat{\mu}_t = \operatorname{med}_{i \in \{1, \dots, n\}} \left\{ x_{t-n+i} - \hat{\beta}_t \cdot (i - n) \right\}$$

Repeated Median (RM) regression

Data-driven
window
width
adaption
for robust
online
moving
window
regression

Matthias
Borowski

Trade-off problem for window width (ww):

- Large $ww \Rightarrow$ smooth signal extraction
- Small $ww \Rightarrow$ exact signal extraction

Approach: data-driven ww selection for RM

At each time point t :

Test: Is window width n adequate for RM fit?

- Yes: estimate signal
- No: decrease n

- **ADORE** – **AD**aptive **O**nline **RE**peated Median
(Schettlinger et al., 2010)
- **SCARM** – **S**lope **C**omparing **A**daptive **R**eppeated **M**edian
(Borowski, 2010)



ADORE test

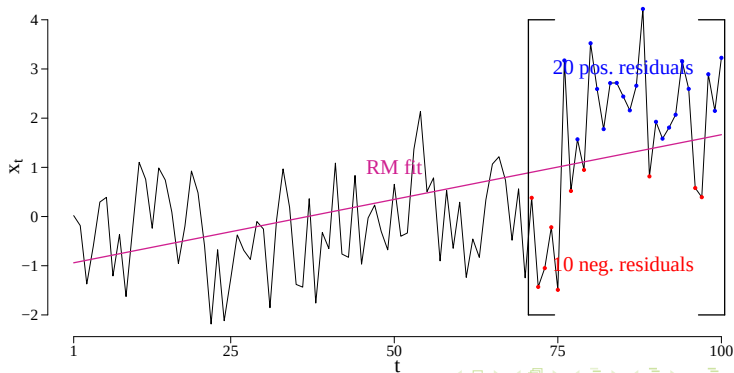
Data-driven
window
width
adaption
for robust
online
moving
window
regression

Matthias
Borowski

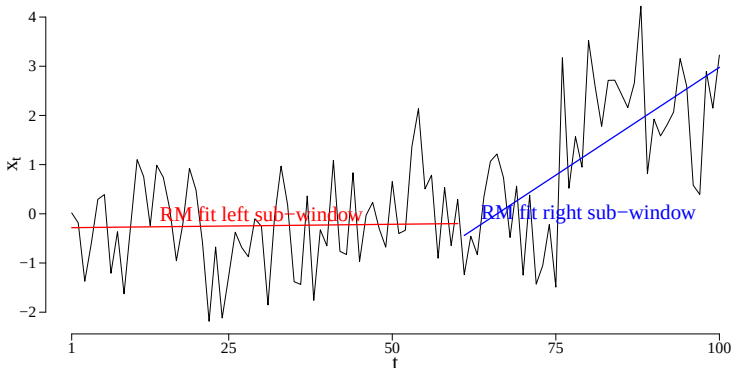
H_0 : ww n adequate, if

neg. residuals $\approx$ # pos. residuals

in right sub-window



$$H_0: \text{ww } n \text{ adequate, iff } \frac{|\hat{\beta}^{\text{left}} - \hat{\beta}^{\text{right}}|}{\sqrt{\text{Var}(\hat{\beta}^{\text{left}} - \hat{\beta}^{\text{right}})}} \text{ is small}$$



$$H_0: \text{ww } n \text{ adequate, iff } \frac{|\hat{\beta}^{\text{left}} - \hat{\beta}^{\text{right}}|}{\sqrt{\text{Var}(\hat{\beta}^{\text{left}} - \hat{\beta}^{\text{right}})}} \text{ is small}$$

Sophisticated estimation of $\text{Var}(\hat{\beta}^{\text{left}} - \hat{\beta}^{\text{right}})$
to increase power:

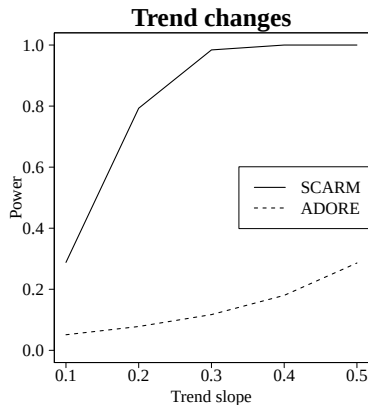
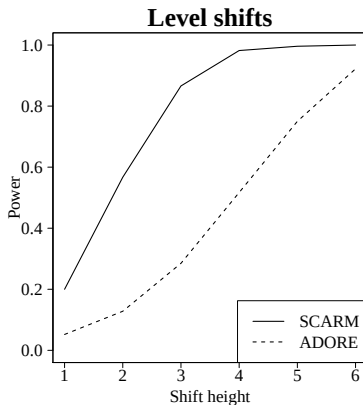
- Use that RM slope is unbiased (symm. noise) and regression equivariant
- Estimate noise variability by regression-free scale estimator (Gelper et al., 2009)

Power comparison ADORE vs. SCARM

Data-driven
window
width
adaption
for robust
online
moving
window
regression

Matthias
Borowski

- Standard normal noise
- Significance level 0.01
- Several sizes of level shifts and trend changes



- ww adaption based on
 - Shift detection (Fried, Gather, 2007)
 - Trend detection (Fried, Imhoff, 2004)
- Further comparisons:
 - Other types of noise
 - Effect of outliers
- SCARM in R package `robfilter`
(Fried, Schettlinger, Borowski, 2010)

- Borowski, M. (2010): Window width adaption for robust moving window regression in online-monitoring time series. *Discussion Paper, to appear*.
- Fried, R., Gather, U. (2007): On rank tests for shift detection in time series. *CSDA* **52**, 221-233.
- Fried, R., Imhoff, M. (2004): On the online detection of monotonic trends in time series. *Biometrical Journal* **46**, 90-102.
- Fried, R., Schettlinger, K., Borowski, M. (2010): robfilter: Robust Time Series Filters. *R package version 2.6.1*
- Siegel, A. F. (1982): Robust regression using repeated medians. *Biometrika* **69**(1), 242-244.
- Gelper, S., Schettlinger, K., Croux, C., Gather, U. (2009): Robust online scale estimation in time series: a regression-free approach. *J. Stat. Plann. Inf.* **139**, 335-349.
- Schettlinger, K., Fried, R., Gather, U. (2010): Real time signal processing by adaptive repeated median filters. *Int. J. Adapt. Control Signal Process.* **24**, 346-362.

Estimation of $\text{Var}(D)$

$$\begin{aligned}\text{Var}(D) &= \text{Var}\left(\hat{\beta}^{\text{left}} - \hat{\beta}^{\text{right}}\right) \\ &\stackrel{*}{=} \text{Var}\left(\hat{\beta}^{\text{left}}\right) + \text{Var}\left(\hat{\beta}^{\text{right}}\right)\end{aligned}$$

*under H_0 for i.i.d. symmetric noise with zero median

$\text{Var}\left(\hat{\beta}\right)$ depends on n and on noise variance σ^2 :

$$\text{Var}\left(\hat{\beta}\right) =: V(n, \sigma) = V(n, 1) \cdot \sigma^2$$

- approximations $\hat{V}(n, 1) =: v_n$ for $n = 5, \dots, 300$ by simulation
- estimate noise variance σ^2 on residuals? $\Rightarrow$ small power!
better: obtain $\hat{\sigma}^2$ by regression-free Q scale estimator
(Gelper et al., 2009)

Estimation of

$$\text{Var}(D) = \text{Var} \left(\hat{\beta}^{\text{left}} \right) + \text{Var} \left(\hat{\beta}^{\text{right}} \right)$$

$$\widehat{\text{Var}(D)} = v_\ell \cdot Q(\mathbf{x}_t^{\text{left}})^2 + v_r \cdot Q(\mathbf{x}_t^{\text{right}})^2$$

$$\rightarrow \text{ if } |d^*| = \left| \frac{d}{\sqrt{v_\ell \cdot Q(\mathbf{x}_t^{\text{left}})^2 + v_r \cdot Q(\mathbf{x}_t^{\text{right}})^2}} \right|$$

is 'too large': reject H_0

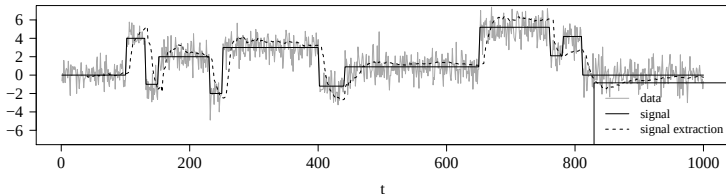
Critical values for D^* ?

- D^* roughly $N(0, 1)$ distributed for several noise distributions
- simulation: quantiles of the emp. distribution of d^* for $N(0, 1)$ noise

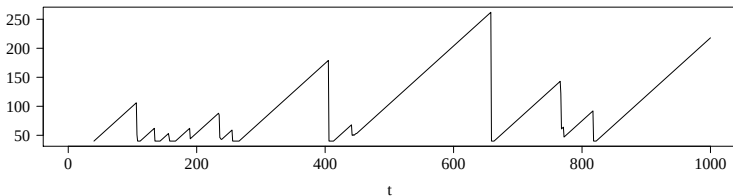
Application: Blocks data

SCARM arguments: $r = 20$, $n_{min} = 40$, $n_{max} = 300$, $\alpha = 0.001$

Blocks Data and SCARM signal extraction



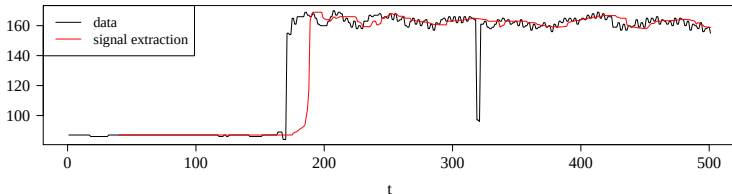
window widths



Application: Real data

SCARM arguments: $r = 40$, $n_{min} = 80$, $n_{max} = 300$, $\alpha = 0.001$

Real data and SCARM signal extraction



window widths

