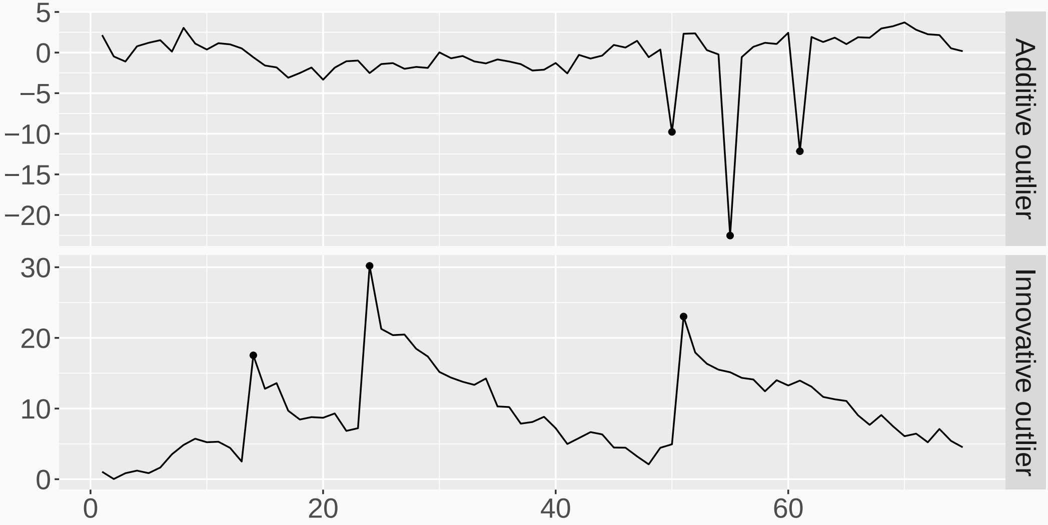
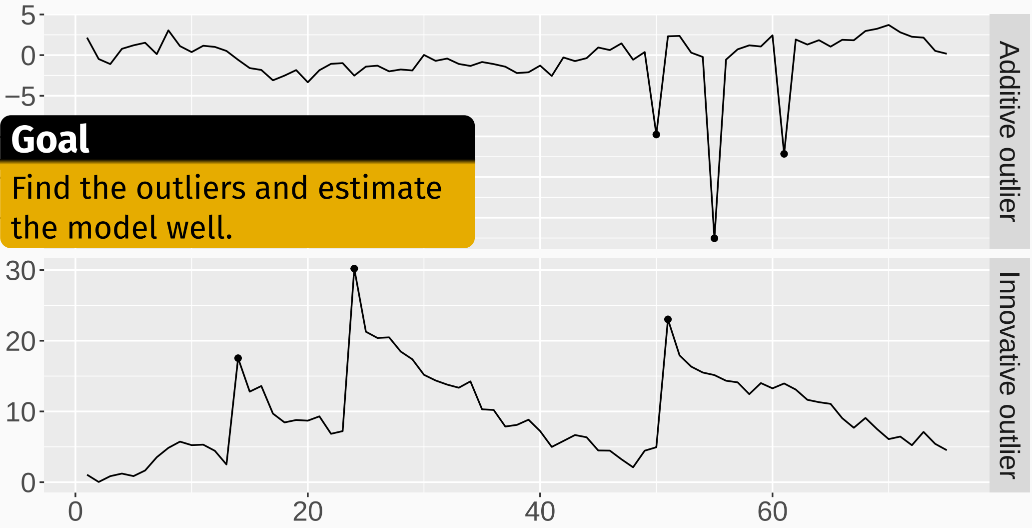
Time series outlier detection using penalised regression

Yangzhuoran Fin Yang
Ines Wilms
Jakob Raymaekers

Outliers in time series



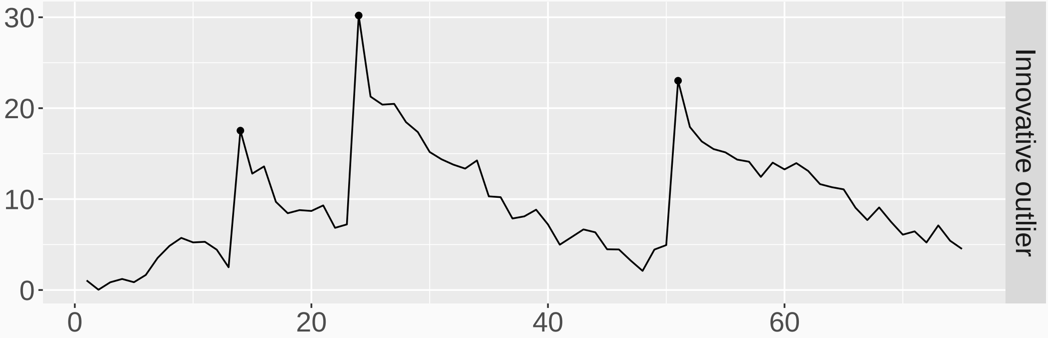
Outliers in time series



Goal

Find the outliers and estimate the model well.

Innovative outlier



Innovative outlier

$$y_t = \gamma_t + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$$

IPOD for Innovative Outliers

Thresholding based Iterative Procedure for Outlier Detection (She and Owen 2011)

Innovative outliers

$$y_t = \gamma_t + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$$

IPOD Objective function

$$\min_{\gamma} f_{\text{soft}}(\varphi, \gamma; \lambda) = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{Y}\varphi - \gamma||_2^2 + \sum_{t=1}^{T} \lambda_t |\gamma_t|.$$

IPOD for Innovative Outliers

Objective function

$$\min_{\gamma} f_{\text{soft}}(\varphi, \gamma; \boldsymbol{\lambda}) = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{Y}\varphi - \gamma||_2^2 + \sum_{t=1}^{T} \lambda_t |\gamma_t|$$

$$\boldsymbol{y} = [y_1, y_2, \ldots, y_T]' \qquad \boldsymbol{Y} = \begin{bmatrix} y_0 & y_{-1} & \cdots & y_{-p+1} \\ y_1 & y_0 & \cdots & y_{-p+2} \\ \vdots & \ddots & & \vdots \\ y_{T-1} & y_{T-2} & \cdots & y_{T-p} \end{bmatrix}$$

IPOD for Innovative Outliers

Objective function

$$\min_{\gamma} f_{\text{soft}}(\varphi, \gamma; \boldsymbol{\lambda}) = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{Y}\varphi - \gamma||_2^2 + \sum_{t=1}^{T} \lambda_t |\gamma_t|$$

Joinly convex in φ and γ - Alternating optimisation

IPOD for Innovative Outliers

Objective function

$$\min_{\gamma} f_{\text{soft}}(\varphi, \gamma; \lambda) = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{Y}\varphi - \gamma||_2^2 + \sum_{t=1}^{T} \lambda_t |\gamma_t|$$

Joinly convex in φ and γ - Alternating optimisation

IPOD

```
while not converged do
```

1. Estimate φ
2. Thresholding $(\boldsymbol{y} - \boldsymbol{Y}\hat{\varphi})$ to obtain optimal γ

```
end while
```

Proximal gradient descent

IPOD

```
while not converged do
```

1. Estimate φ
2. Thresholding $(\boldsymbol{y} - \boldsymbol{Y}\hat{\varphi})$ to obtain optimal γ

```
end while
```

Thresholding

$$r_t = y_t - \varphi y_{t-1} \qquad \Theta_{soft}(r;\lambda) = \begin{cases} r - \lambda & r > \lambda \\ 0 & -\lambda \le r \le \lambda \\ r + \lambda & r < \lambda \end{cases}$$

Proximal gradient descent

IPOD

```
while not converged do
```

1. Estimate φ
2. Thresholding $(\boldsymbol{y} - \boldsymbol{Y}\hat{\varphi})$ to obtain optimal γ

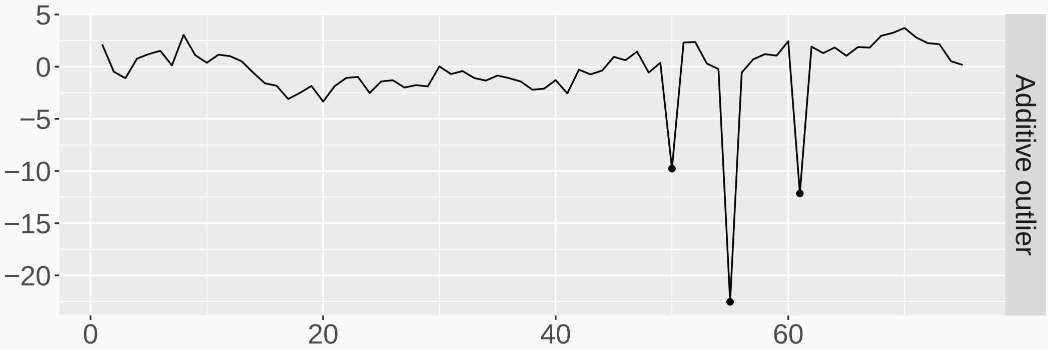
```
end while
```

Same for cross-sectional data

Thresholding

$$r_t = y_t - \varphi y_{t-1} \qquad \Theta_{soft}(r;\lambda) = \begin{cases} r - \lambda & r > \lambda \\ 0 & -\lambda \leq r \leq \lambda \\ r + \lambda & r < \lambda \end{cases}$$

Additive outlier



Additive outlier

$$y_t = \gamma_t + u_t$$

$$u_t = \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \cdots + \varphi_p u_{t-p} + \varepsilon_t$$

IPOD for Additive Outliers?

Additive outliers

$$y_t = \gamma_t + u_t$$

$$u_t = \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \cdots + \varphi_p u_{t-p} + \varepsilon_t$$

IPOD Objective function

$$\min_{\gamma} f_{\text{soft}}(\varphi, \gamma; \lambda) = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{U}\varphi - \gamma||_2^2 + \sum_{t=1}^{T} \lambda_t |\gamma_t|.$$

IPOD for Additive Outliers?

Objective function

$$\min_{\gamma} f_{\text{soft}}(\varphi, \gamma; \lambda) = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{U}\varphi - \gamma||_2^2 + \sum_{t=1}^{T} \lambda_t|\gamma_t|$$

IPOD?

```
while not converged do
```

1. Estimate φ
2. Thresholding $(\boldsymbol{y} - \boldsymbol{U}\hat{\varphi})$ to obtain optimal γ

```
end while
```

Proximal gradient descent?

IPOD?

```
while not converged do
```

1. Estimate φ
2. Thresholding $(\boldsymbol{y} - \boldsymbol{U}\hat{\varphi})$ to obtain optimal γ

```
end while
```

Thresholding

$$r = y - \hat{u} \qquad \Theta_{soft}(r; \lambda) = \begin{cases} r - \lambda & r > \lambda \\ 0 & -\lambda \leq r \leq \lambda \\ r + \lambda & r < \lambda \end{cases}$$

Proximal gradient descent?

IPOD?

```
while not converged do
```

1. Estimate φ
2. Thresholding $(\boldsymbol{y} - \boldsymbol{U}\hat{\varphi})$ to obtain optimal γ

```
end while
```

Thresholding

No longer solves a consistent objective function!

$$r = y - \hat{u} \qquad \Theta_{soft}(r;\lambda) = \begin{cases} 0 & -\lambda \leq r \leq \lambda \\ r + \lambda & r < \lambda \end{cases}$$

Solving it properly

After some derivation...

For given φ, solving for γ can be expressed as a lasso problem.

$$\min_{\gamma} f_{\text{soft}}(\gamma; \boldsymbol{\lambda}, \varphi) = \frac{1}{2}||\boldsymbol{\Phi}^* \gamma - \boldsymbol{\delta}||_2^2 + \sum_{t=1}^{T} \lambda_t |\gamma_t|$$

`while not converged do`

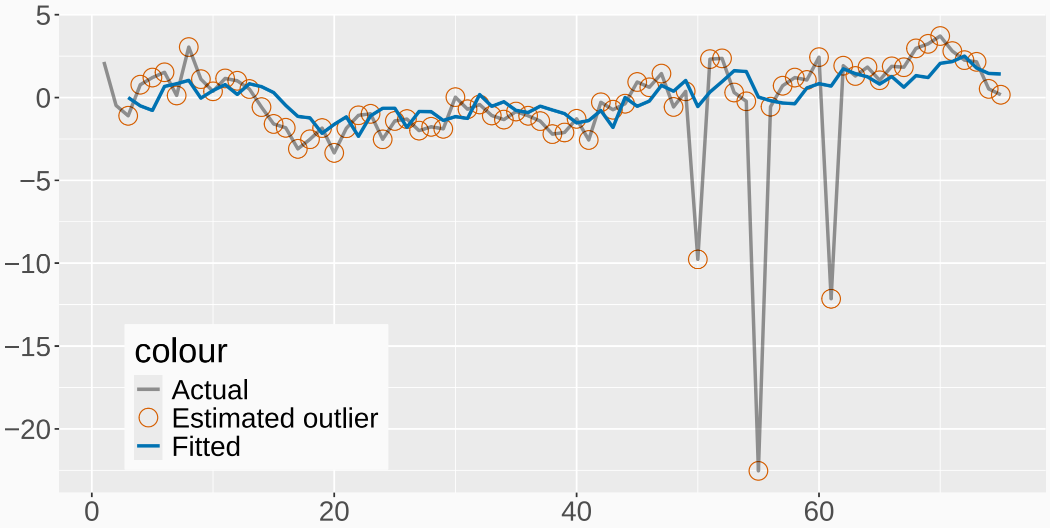
1. Estimate φ
2. Inner proximal gradient descent loop
   - `while not converged do`

     Thresholding to obtain optimal γ

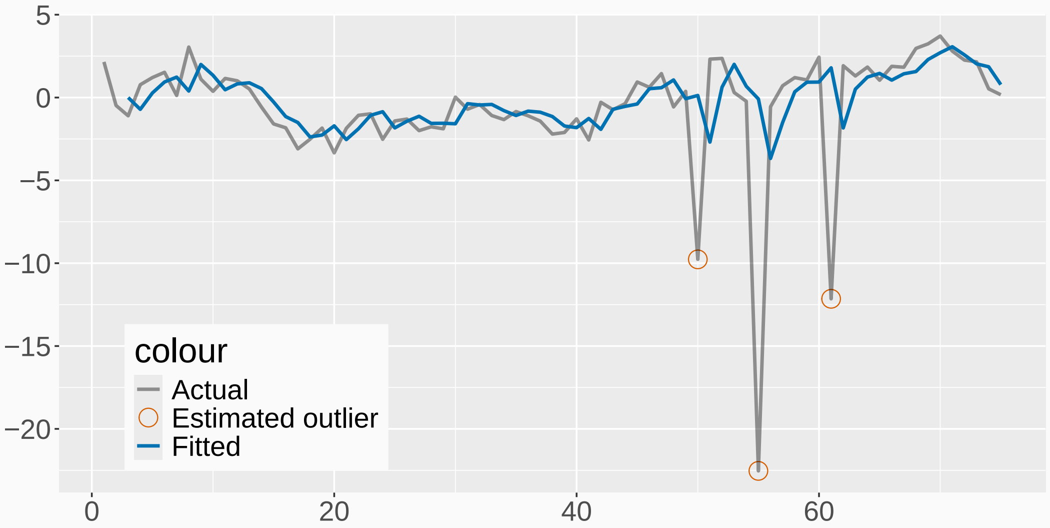
     `end while`

`end while`

Optimisation: Iteration 0



Optimisation: Iteration 1



Observation

Outliers can be identified

Observation

Outliers can be identified but **underestimated**.

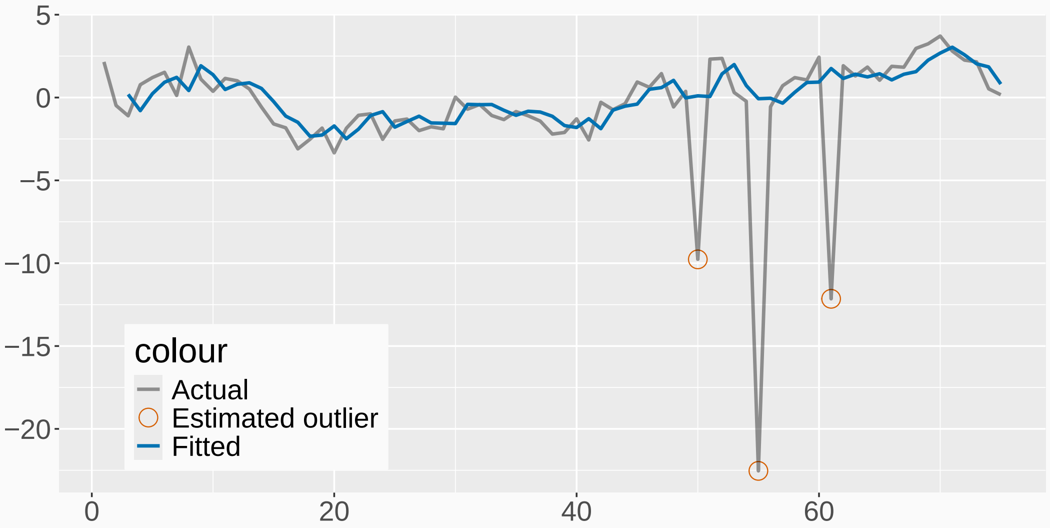
Observation

Outliers can be identified but **underestimated**.

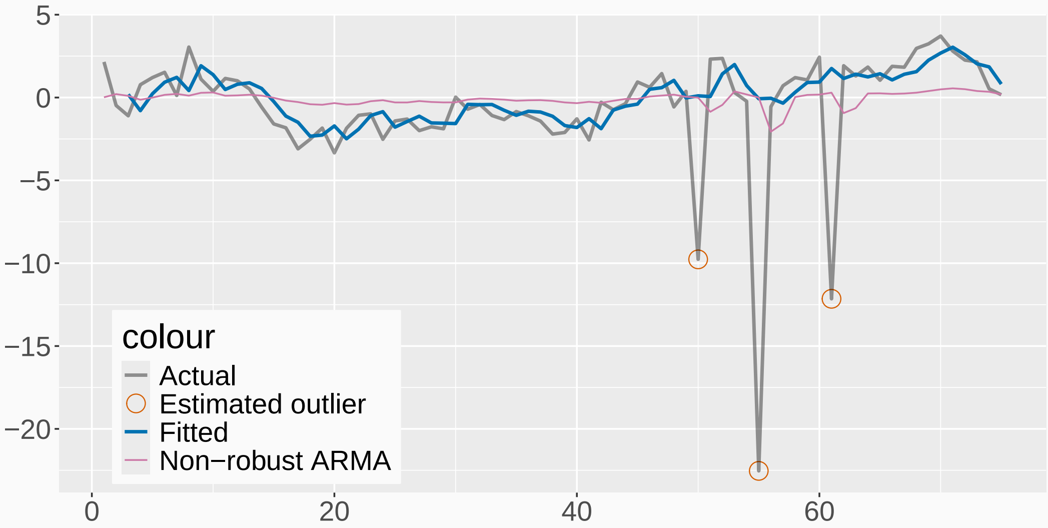
Reweighting

1. Remove the effects of estimated outlier and refit the model
2. Keep $(\boldsymbol{y} - \boldsymbol{U}\hat{\varphi})$ at the locations of the identified outlier as the new estimated outliers
3. Refit the model with the effects of the new outliers removed

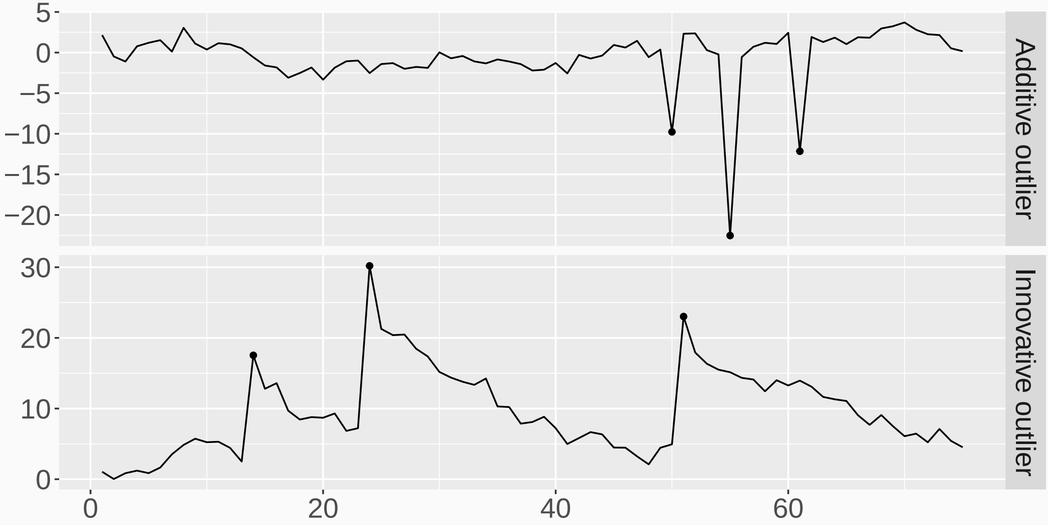
Iteration… and Reweight



Non-robust ARMA



Outliers in time series



AR(1): First obs is an outlier

Objective function

$$\min \sum_{t=2}^{T} \tilde{e}_t^2$$

$$\tilde{e}_2 = y_2 - \gamma_2 - \varphi(y_1 - \gamma_1) = y_2 - \varphi(y_1 - \gamma_1) = e_2 + \varphi\gamma_1$$

$$\tilde{e}_3 = y_3 - \gamma_3 - \varphi(y_2 - \gamma_2) = y_3 - \varphi y_2 = e_3$$

$$\tilde{e}_t = e_t, \text{ where } e_t = y_t - \varphi y_{t-1}$$

Optimal γ

$$\gamma_1^* = \frac{e_2}{\varphi}$$

AR(1): First two obs are outliers

Objective function

$$\min \sum_{t=2}^{T} \tilde{e}_t^2$$

$$\tilde{e}_2 = y_2 - \gamma_2 - \varphi(y_1 - \gamma_1) = y_2 - \gamma_2 - \varphi(e_1 - \gamma_1) = e_2 + \varphi\gamma_1 - \gamma_2$$

$$\tilde{e}_3 = y_3 - \gamma_3 - \varphi(y_2 - \gamma_2) = y_3 - \varphi(y_2 - \gamma_2) = e_3 + \varphi\gamma_2$$

$$\tilde{e}_t = e_t$$

Optimal γ

$$\gamma_2^* = -\frac{e_3}{\varphi} = -\frac{y_3 - \varphi y_2}{\varphi} \qquad \gamma_1^* = -\frac{e_2 - \gamma_2^*}{\varphi} = -\frac{e_2}{\varphi} - \frac{e_3}{\varphi^2}$$

AR(1): First two obs are outliers

Objective function

$$\min \sum_{t=2}^{T} \tilde{e}_t^2$$

$$\tilde{e}_2 = y_2 - \gamma_2 - \varphi(y_1 - \gamma_1) = y_2 - \gamma_2 - \varphi(e_1 - \gamma_1) = e_2 + \varphi\gamma_1 - \gamma_2$$

$$\tilde{e}_3 = y_3 - \gamma_3 - \varphi(y_2 - \gamma_2) = y_3 - \varphi(y_2 - \gamma_2) = e_3 + \varphi\gamma_2$$

Optimal value of γ_i depends on optimal value of γ_j

Optimal γ

$$\gamma_2^* = -\frac{e_3}{\varphi} = -\frac{y_3 - \varphi y_2}{\varphi} \qquad \gamma_1^* = -\frac{e_2 - \gamma_2^*}{\varphi} = -\frac{e_2}{\varphi} - \frac{e_3}{\varphi^2}$$

AR(1): First two obs are outliers

Objective function

$$\min \sum_{t=2}^{T} \tilde{e}_t^2$$

$$\tilde{e}_2 = y_2 - \gamma_2 - \varphi(y_1 - \gamma_1) = y_2 - \gamma_2 - \varphi(e_1 - \gamma_1) = e_2 + \varphi\gamma_1 - \gamma_2$$

$$\tilde{e}_3 = y_3 - \gamma_3 - \varphi(y_2 - \gamma_2) = y_3 - \varphi(y_2 - \gamma_2) = e_3 + \varphi\gamma_2$$

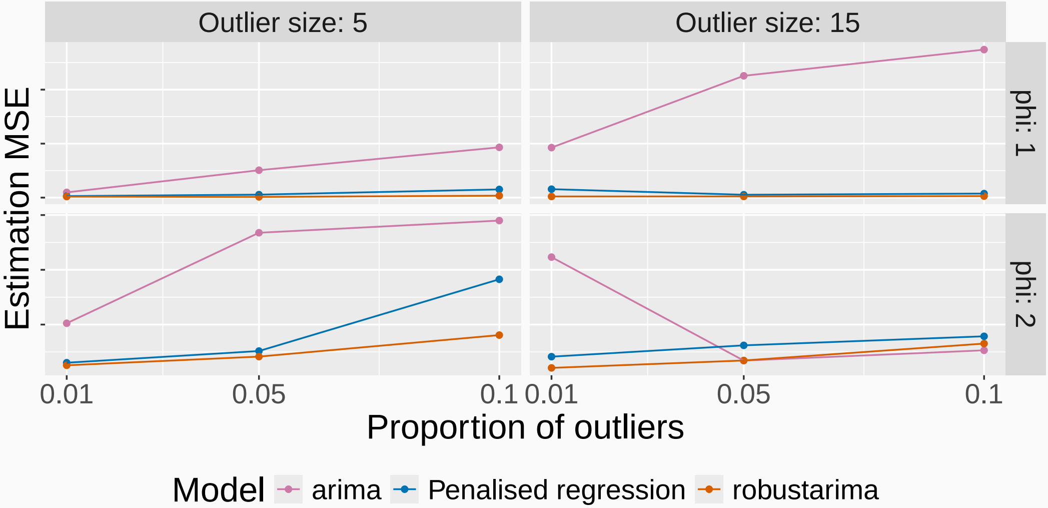
Optimal value of γ_i depends on optimal value of γ_j

Optimal γ

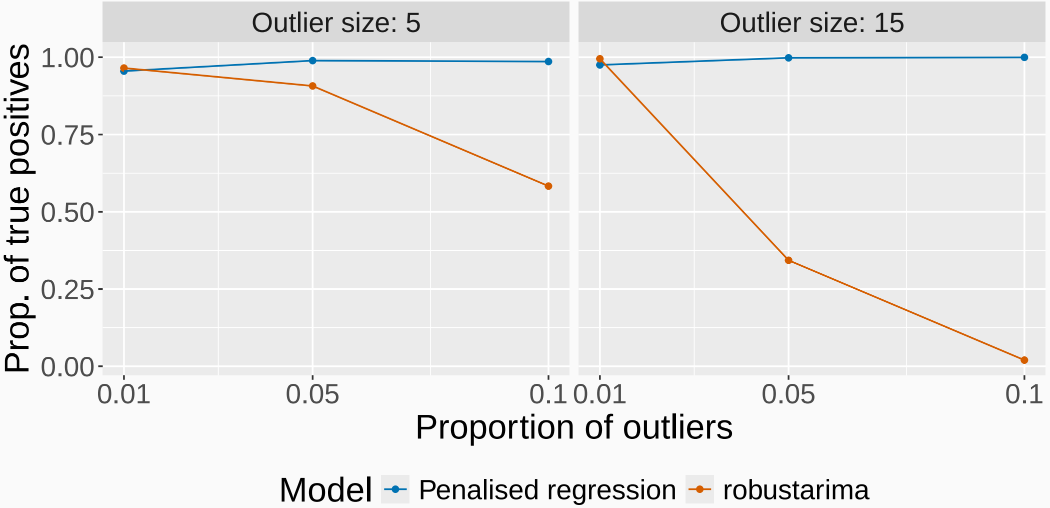
$$\gamma_2^* = -\frac{e_3}{\varphi} = -\frac{y_3}{\varphi} \quad \ldots \quad \frac{\ldots_3}{\varphi^2}$$

Never the case for innovative outliers!

Performance: Model estimation (no reweight)



Performance: Outlier identification



Contact

Slides:
yangzhuoranyang.com/
talk/isf-2026

yangzhuoranyang.com

yangzhuoran.yang
@maastrichtuniversity.nl

References

She, Yiyuan, and Art B Owen. 2011. "Outlier detection using nonconvex penalized regression." *J. Am. Stat. Assoc.* 106 (494): 626–39. https://doi.org/10.1198/jasa.2011.tm10390.