Fast and Robust Least Squares Estimation in Corrupted Linear Models

Brian McWilliams*, Gabriel Krummenacher*, Mario Lucic, Joachim M. Buhmann

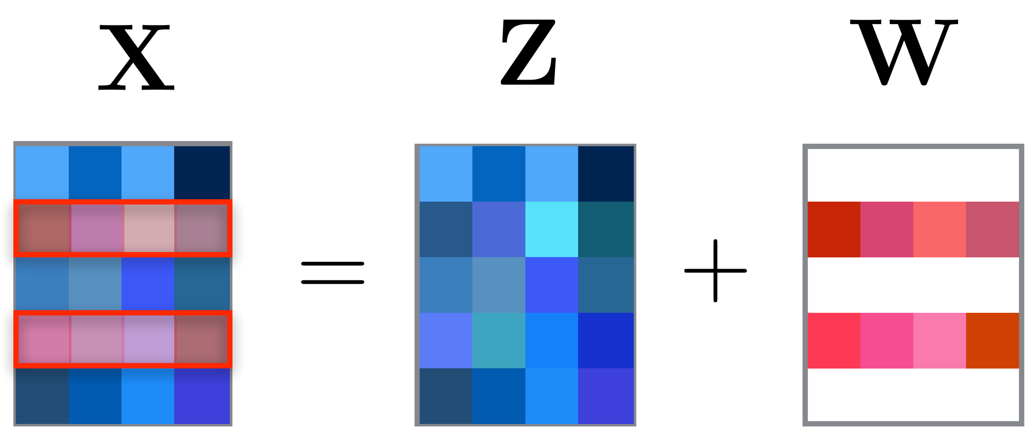
ETH zürich

Statistical Model with Corruptions

We consider a variant of the standard linear model:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \epsilon$$

But we observe $\mathbf{X} = \mathbf{Z} + \mathbf{W}$.
Each row of $\mathbf{W}$ is non-zero with probability π and 0 otherwise.



- Accounts for more realistic setting where measurements aren't perfect.
- Least squares solution is biased $\rightarrow$ Randomized approximations are also biased.
- Corrupted points are "outliers" but the SRHT doesn't help.
- We need a more sophisticated sense of what an outlier is.
- Theoretical Results: $\mathbf{Z}$, $\mathbf{W}$, ϵ are sub-Gaussian random variables.

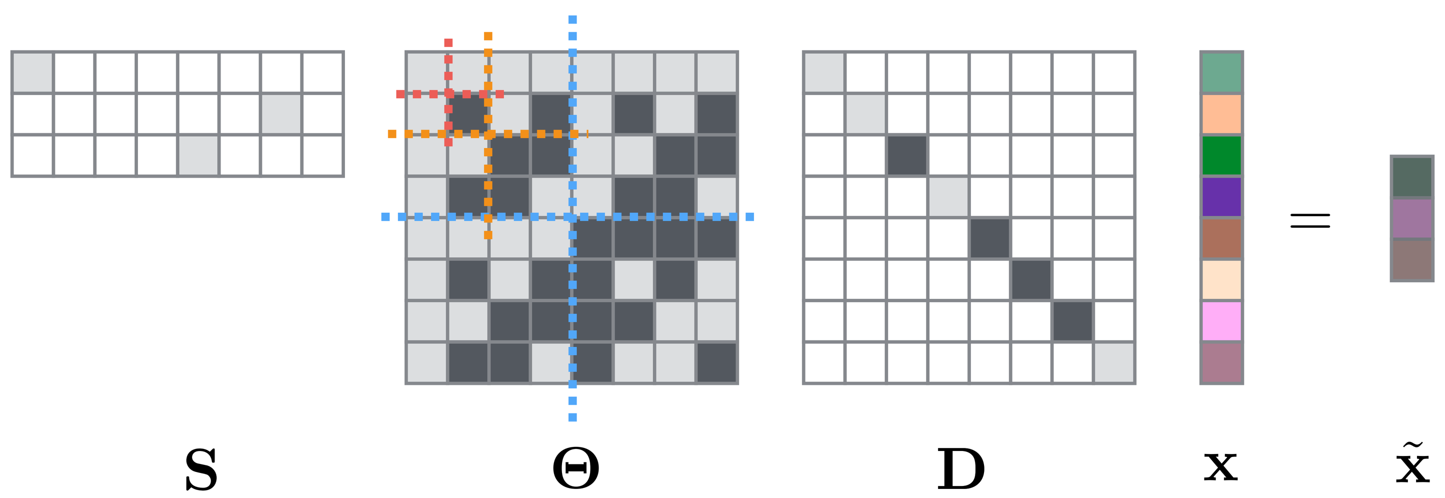
Subsampled Randomized Hadamard Transform

$$\tilde{\mathbf{X}}_S \in \mathbb{R}^{n_{subs}\times p} = \sqrt{\frac{n}{n_{subs}}}\mathbf{SHD}\cdot\mathbf{X}$$

- $\mathbf{S}$ is a $n_{subs} \times n$ subsampling matrix.
- $\mathbf{D}$ is a diagonal with n entries drawn from $\{-1, 1\}$.
- $\mathbf{H} \in \mathbb{R}^{n\times n}$ is a normalized Walsh-Hadamard matrix which is defined recursively as

$$\mathbf{H}_n = \begin{bmatrix} \mathbf{H}_{n/2} & \mathbf{H}_{n/2} \\ \mathbf{H}_{n/2} & -\mathbf{H}_{n/2} \end{bmatrix}, \quad \mathbf{H}_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}.$$

  We set $\mathbf{H} = \frac{1}{\sqrt{n}}\mathbf{H}_n$ so it has orthonormal columns.
- Approximately uniformizes leverage scores.



Projected and Subsampled LS (SRHT-LS)

1. Project and subsample using SRHT:
$$(\tilde{\mathbf{X}}_S, \tilde{\mathbf{y}}_S) = \mathbf{SHD}\cdot(\mathbf{X}, \mathbf{y})$$
2. Solve
$$\tilde{\boldsymbol{\beta}}_{SRHT} = \arg\min_{\boldsymbol{\beta}} \|\tilde{\mathbf{y}}_S - \tilde{\mathbf{X}}_S\boldsymbol{\beta}\|^2$$

- Cost of SRHT is $O(n \log n_{subs})$.
- $\tilde{\boldsymbol{\beta}}_{SRHT}$ is a good approximation to $\widehat{\boldsymbol{\beta}}$.

Influence [Belsley, Kuh & Welsch, 1981]

Leave out the point $(\mathbf{x}_i, y_i)$:

$$\|\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{-i}\|^2 = \frac{e_i^2 \cdot l_i}{(1-l_i)^2}$$

Where

- $e_i = y_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}$ is the residual error of data point i.
- $l_i = \mathbf{x}_i(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i^\top$ is the statistical leverage (outlyingness) of point i.

Influence Weighted (IWS-LS)

1: ***Solve*** $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$

2: ***Compute influence*** d_i **for all points.**

3: ***Sample rows*** $(\tilde{\mathbf{X}}, \tilde{\mathbf{y}})$ ***of*** $(\mathbf{X}, \mathbf{y})$ ***inversely proportional to*** d_i

4: ***Solve*** $\widehat{\boldsymbol{\beta}}_{\mathrm{IWS}} = \arg\min_{\boldsymbol{\beta}} \|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}\|^2$

- Influence can be approximated cheaply using $\tilde{\boldsymbol{\beta}}_{SRHT}$.
- $\tilde{\boldsymbol{\beta}}_{IWS}$ is
  - a good approximation to $\widehat{\boldsymbol{\beta}}$ when there are no corruptions.
  - a good estimator of $\boldsymbol{\beta}$ when there are corruptions.

Residual Weighted (aRWS-LS)

1: ***Solve*** $\tilde{\boldsymbol{\beta}}_{SRHT} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{SHD}\cdot(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|^2$

2: ***Estimate residuals:*** $\tilde{\mathbf{e}} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{SRHT}$

3: ***Sample rows*** $(\tilde{\mathbf{X}}, \tilde{\mathbf{y}})$ ***of*** $(\mathbf{X}, \mathbf{y})$ ***inversely proportional to*** $\tilde{e}_i^2$

4: ***Solve*** $\tilde{\boldsymbol{\beta}}_{RWS} = \arg\min_{\boldsymbol{\beta}} \|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}\|^2$

Influence sampling; Corruptions

For $n \gtrsim \frac{\sigma_x^2\sigma_w^2}{\lambda_{\min}(\Sigma_{\Theta x})} p \log p$ we have

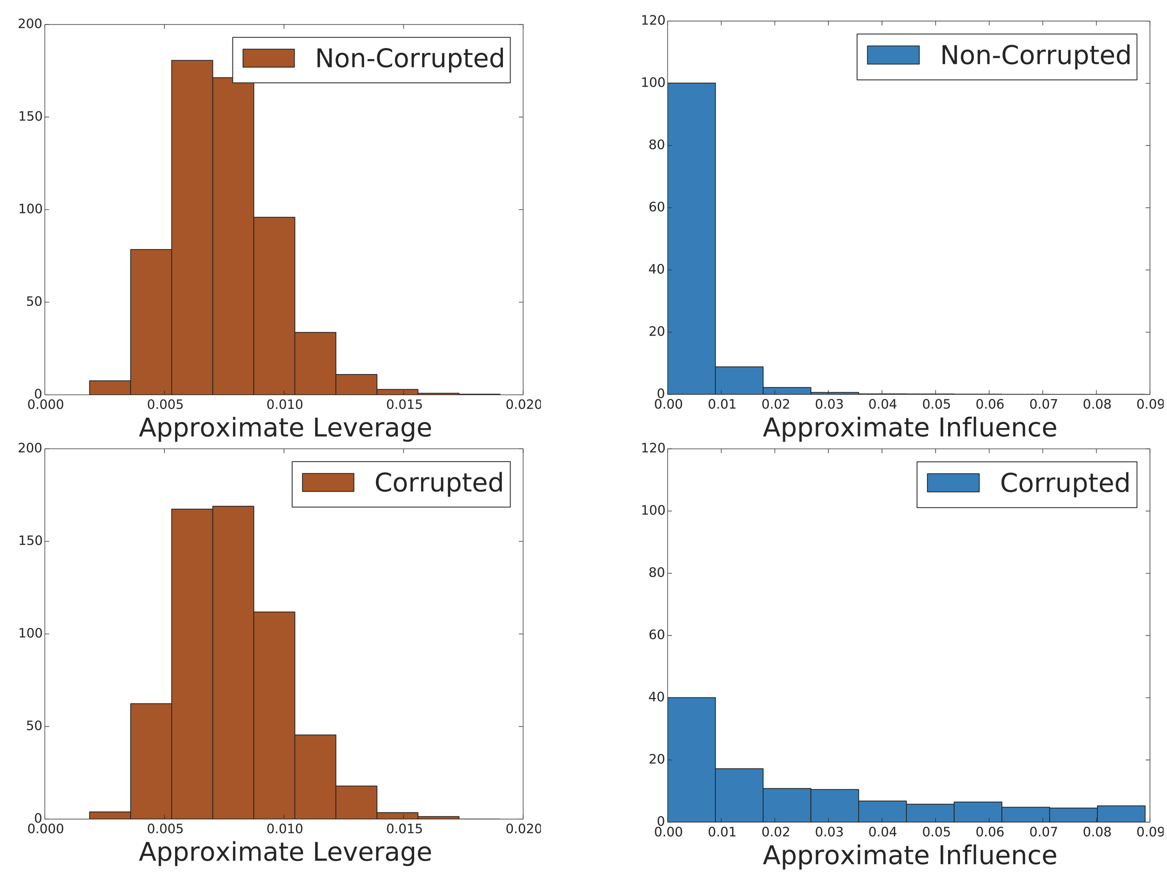
$$\|\widehat{\boldsymbol{\beta}}_{\mathrm{IWS}} - \boldsymbol{\beta}\| \lesssim \left(\left(\sigma_\epsilon\sigma_x + \frac{\pi\sigma_\epsilon}{(\sigma_w+1)} + \pi\|\boldsymbol{\beta}\|\right)\sqrt{\frac{p\log p}{n_{subs}}} + \pi\sqrt{p}\|\boldsymbol{\beta}\|\right)\cdot\frac{1}{\lambda}$$

where $0 < \lambda \leq \lambda_{\min}(\Sigma_{\Theta x})$ and $\Sigma_{\Theta x}$ is the covariance of the influence weighted subsampled data.

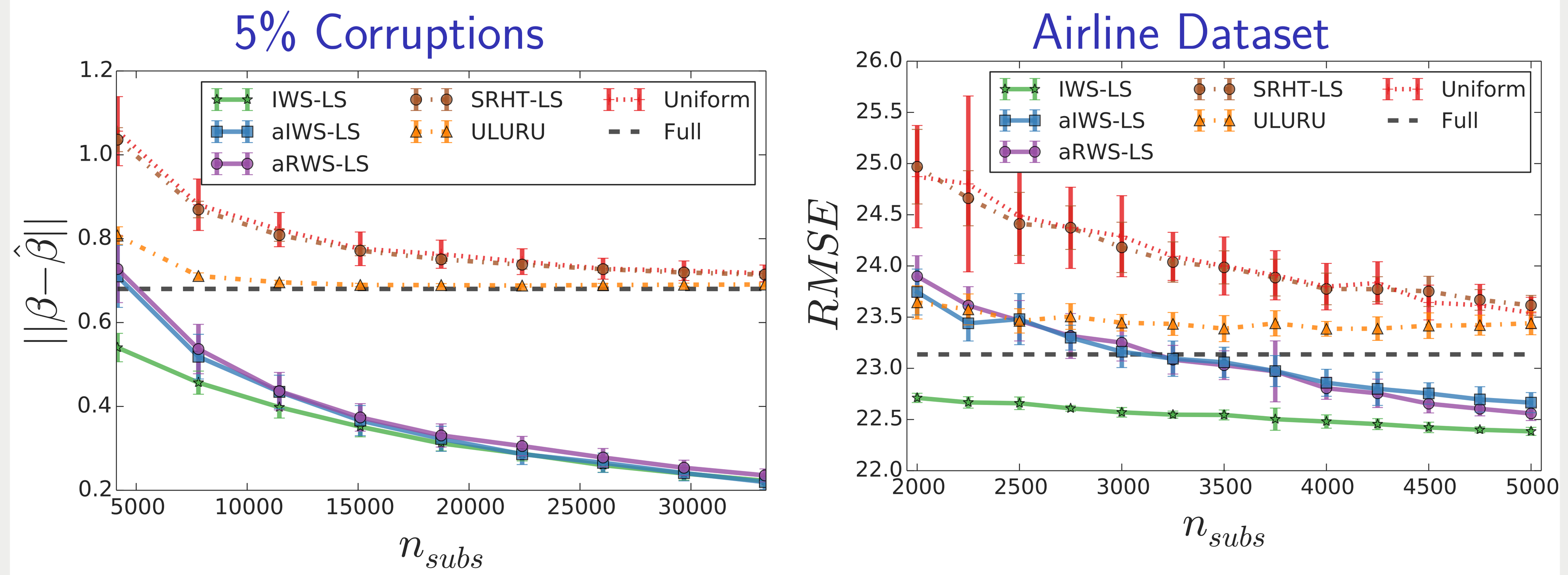
Influence vs Leverage

Non-Corrupted

Corrupted



Empirical Results: Corrupted Setting



Software & Paper

Large Scale Randomized Regression (LRR) package
`http://people.inf.ethz.ch/kgabriel/software.html`

McWilliams, Krummenacher, Lucic & Buhmann
"Fast and Robust Least Squares Estimation in Corrupted Linear Models." `http://arxiv.org/abs/1406.3175`