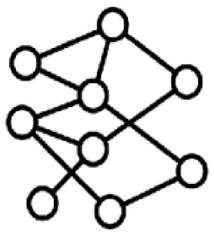
Robustness of OLS to Sample Removals

Eyar Azar* Michael Feldman† Boaz Nadler*

* Weizmann Institute of Science, † New York University

Learning Workflow

Data → **Training Model** → **Output**

$(\boldsymbol{x}_i, y_i)$ →  → $\boldsymbol{f}_{\boldsymbol{\theta}}(\cdot)$

Can you trust this model?

Model Stability Against Data Perturbation

We investigated model stability against perturbations in the training data.

Specifically, we analyzed how a small fraction of samples impacts the learned model.

This area of research is known as ***robustness auditing***.

Recent work [1] demonstrates that removing **< 0.5%** of samples can be enough to completely overturn model conclusions.

Robustness Auditing of OLS

- Labeled Data: $(\boldsymbol{x}_i, y_i)_{\boldsymbol{i}\in[\boldsymbol{n}]}$
- Full OLS estimator:

$$\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{y}$$

- For subset $S \subset [n]$, define the partial OLS estimator:

$$\widehat{\boldsymbol{\beta}}^S = (X_S^\top X_S)^{-1} X_S^\top \boldsymbol{y}_S$$

Robustness metric:

For fixed direction $\boldsymbol{v} \in \mathbb{S}^{p-1}$:

$$\Delta_k(\boldsymbol{v}) = \max_{|S|\geq n-k} \langle \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^S, \boldsymbol{v} \rangle$$

Example:

$\Delta_k(\boldsymbol{e}_1)$ quantifies the sensitivity of $\hat{\beta}_1$ to k removals.

If $\hat{\beta}_1 > 0$ and $\Delta_k(\boldsymbol{e}_1) > |\hat{\beta}_1|$, then

$$\operatorname{sgn} \hat{\beta}_1^S \neq \operatorname{sgn} \hat{\beta}_1$$

Practical implication:

If $x^{(1)}$ (first coordinate of $\boldsymbol{x}$) is a treatment indicator, flipping the sign of $\hat{\beta}_1$ reverses the estimated treatment effect. This fragility renders the study's conclusions completely unreliable.

Research Question

How robust is OLS to sample removals across different regimes of dimension $\boldsymbol{p}$ and deletion size $\boldsymbol{k}$?

Theoretical Results

Note: The optimal predictor $E[y \mid \boldsymbol{x}]$ might be non-linear, making the linear OLS fit fundamentally misspecified.
Thus, we evaluate robustness under model misspecification.

General Misspecified Model:

- $\boldsymbol{x}$ and y are zero mean sub-Gaussians

Theorem 1

If $k \ll n$, with high probability

$$\max_S ||\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^S|| \leq O(k/n \log n/k)$$

Key takeaway: If $k \ll n$ OLS maintains robustness regardless of the dimension, even without a specified linear relationship.

Linear Gaussian Model:

- $y = \boldsymbol{\beta}^\top \boldsymbol{x} + \varepsilon; \;\; \boldsymbol{x} \sim N(0, \Sigma)$
- ε is zero mean sub-Gaussian

Theorem 2

Let $\alpha = k/n$, with high probability

$$\Delta_k(\boldsymbol{v}) \geq \frac{1}{1-\alpha} ||\Sigma^{-1}\boldsymbol{v}||\, E[\varepsilon z 1(\varepsilon z \leq q_\alpha)]$$

$z \sim N(0,1), q_\alpha$ is the α-quantile of εz

Key takeaway: When the removal fraction α is large ($k \propto n$), OLS is fundamentally *non-robust*.

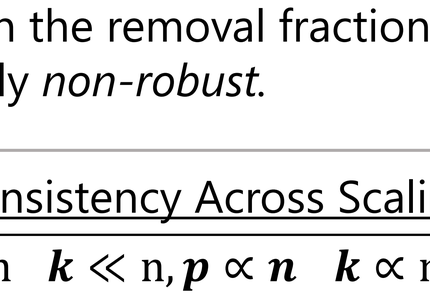
Robustness and Consistency Across Scaling Regimes:

Region	$\boldsymbol{k}, \boldsymbol{p} \ll$ n	$\boldsymbol{k} \ll$ n, $\boldsymbol{p} \propto \boldsymbol{n}$	$\boldsymbol{k} \propto$ n, $\boldsymbol{p} \ll \boldsymbol{n}$	$\boldsymbol{k}, \boldsymbol{p} \propto \boldsymbol{n}$
Robust	✓	✓	✗	✗
Consistent	✓	✗	✓	✗

Huber Regression

For heavy-tailed responses, OLS is highly sensitive to sample removals.

This sensitivity can be attenuated by regression with a *Huber loss*.



OLS vs. Huber on the Cash Transfer dataset [2]

$\boldsymbol{n}$	$> 5\sigma_y$	$\hat{\beta}_1^{OLS}$	$\boldsymbol{k}_{\text{sign}}^{\boldsymbol{OLS}}$	$\hat{\beta}_1^{Huber}$	$\boldsymbol{k}_{\text{sign}}^{\boldsymbol{Huber}}$
10781	48	16.5	224	14.2	570
9489	48	28.7	314	22.1	817
3769	20	23.2	**21**	11.2	**124**
10368	56	32.5	555	26	1145
4191	19	21.1	**26**	13.4	**162**

[1] Tamara Broderick, Ryan Giordano, and Rachael Meager (2020). An Automatic Finite-Sample Robustness Metric: When Can Dropping a Little Data Make a Big Difference?
[2] Angelucci, M. and De Giorgi, G. (2009). Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption?