Residuals Scaling for Bootstrap Prediction Interval in Regression

Sha Taole
Bachelor of Science (Statistics)

Research Colloquium for Science Undergraduate Students 2023-24
Poster No.: A1
Name: Sha Taole
University No.: 3035770898
Student's Major: Statistics
Supervisor: Prof. Stephen Lee

Problem Statement

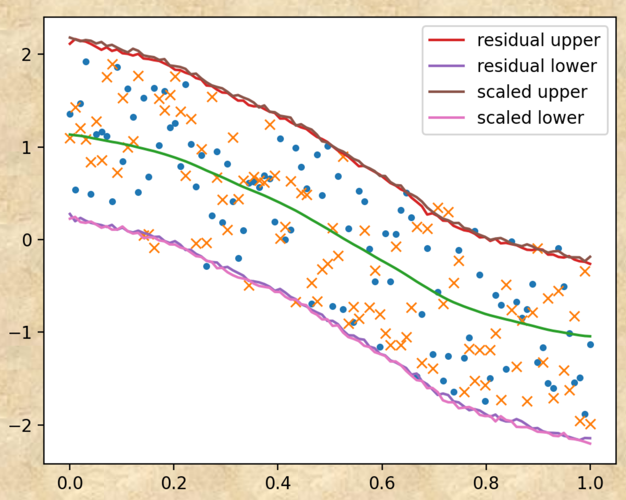
i.i.d. data pairs $(\boldsymbol{x}_i,\ y_i)$ $i = 1,2,\dots,N$ of (X,Y)
Modelling assumption: $y = f(\boldsymbol{x}) + \epsilon$
Unknown function: $f(\boldsymbol{x}) = E(Y|X=\boldsymbol{x})$
Noise: $\epsilon \sim F$, mean 0, variance σ^2
Regression estimator at a future input: $\hat{f}(\boldsymbol{x}^*) = \hat{E}(Y|X=\boldsymbol{x}^*)$
Target prediction interval: $I(\boldsymbol{x}^*) = [l(\boldsymbol{x}^*), u(\boldsymbol{x}^*)]$
β-expectation: $\Pr\left(y^* \in I(\boldsymbol{x}^*)\right) = \beta$

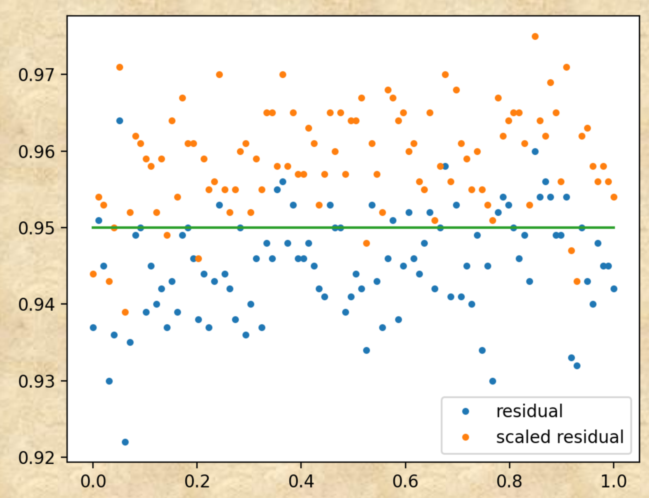
Handling the Heteroscedasticity

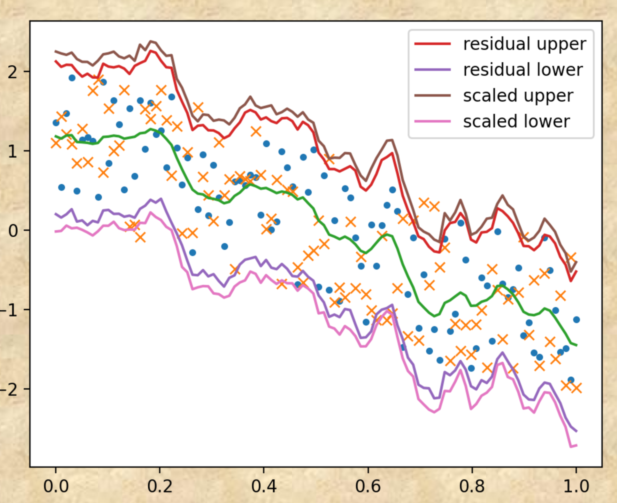
Alternative modelling assumption: $y = f(\boldsymbol{x}) + \epsilon, \epsilon = h(\boldsymbol{x})\epsilon'$
Noise: $\epsilon' \sim F$, mean 0, variance σ^2
Heteroscedasticity scale: $h(\boldsymbol{x})$

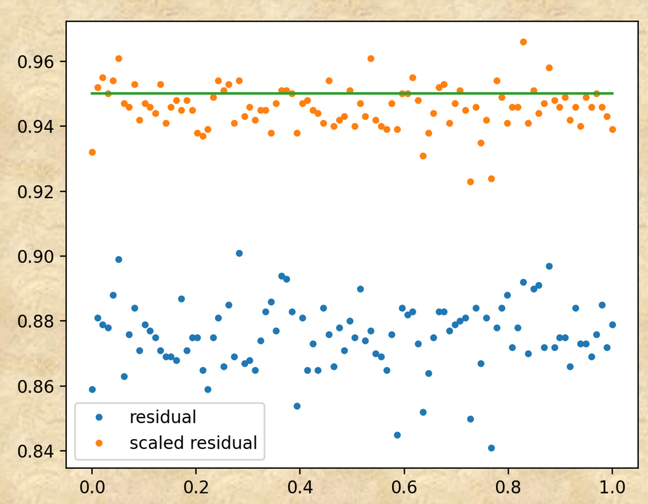
- $h(\boldsymbol{x}) = 1$ backs to the homoscedastic situation
- $h(\boldsymbol{x}) = E(|\epsilon| \mid X = \boldsymbol{x})$ forwards to the heteroscedastic situation

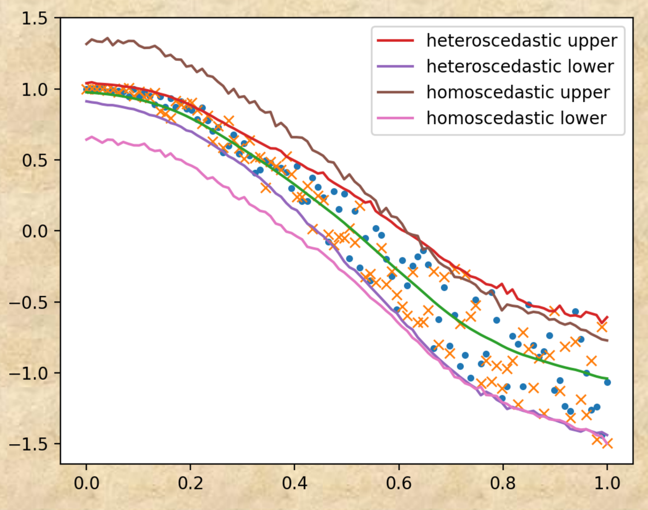
Regressing $\hat{\epsilon}_i / s_i\hat{\epsilon}_i$ on $\boldsymbol{X}$ to have the $\hat{h}(\boldsymbol{x})$ (Lei et al., 2018)

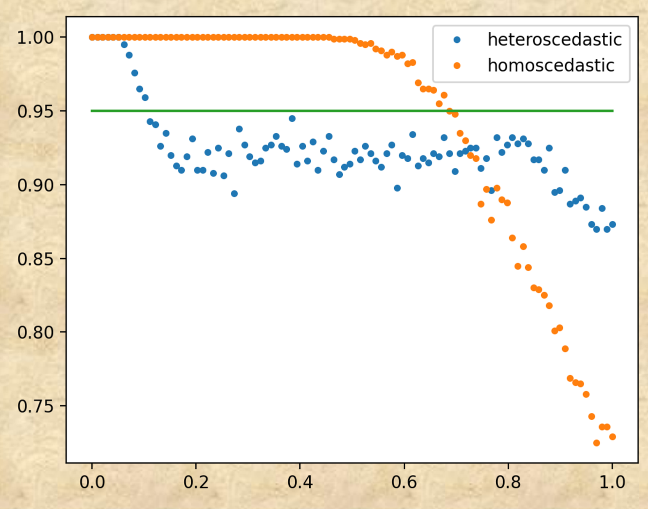












Bootstrap Prediction Interval

Decompose the output: $y^* = \hat{f}(\boldsymbol{x}^*) + f(\boldsymbol{x}^*) - \hat{f}(\boldsymbol{x}^*) + \epsilon^*$
The unknown target: $(R^* = f(\boldsymbol{x}^*) - \hat{f}(\boldsymbol{x}^*) + \epsilon^*) \sim G^*$
Approximation: $\left(\widetilde{R_b^*} = \hat{f}(\boldsymbol{x}^*) - \widehat{f_b}(\boldsymbol{x}^*) + \widetilde{\epsilon_b^*}\right)$, *empirically* $\widetilde{G^*}$
b denotes b^{th} Bootstrap replicate
Bootstrap prediction interval:

$$I^{(B)}(\boldsymbol{x}^*) = \left[\hat{f}(\boldsymbol{x}^*) + \widetilde{G^*}^{-1}\left(\frac{1-\beta}{2}\right), \hat{f}(\boldsymbol{x}^*) + \widetilde{G^*}^{-1}\left(\frac{1-\beta}{2}\right)\right]$$

The idea back to Stine (1985) in a linear regression setting

Generate Bootstrap Replicates

Sample residuals: $\widehat{\epsilon_i} = y_i - \hat{f}(\boldsymbol{x}_i)$, *empirically* $\widehat{F_N}$
Generate $y_{b,i} = \hat{f}(\boldsymbol{x}_i) + \epsilon_i,\ \epsilon_i \sim \widehat{F_N}$
$\widehat{f_b}(\boldsymbol{x}^*)$ is dependent on $(\boldsymbol{x}_i,\ y_{b,i})$ $i = 1,2,\dots,N$
Noise term: $\widetilde{\epsilon_b^*} \sim \widehat{F_N}$

Asymptotic Validity

Assumption: as $N \to \infty$, $\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x}) \xrightarrow{p} 0$, $\widehat{f_b}(\boldsymbol{x}) - \hat{f}(\boldsymbol{x}) \xrightarrow{p} 0$
Asymptotic β-expectation: $\Pr\left(y^* \in I^{(B)}(\boldsymbol{x}^*)\right) \xrightarrow{p} \beta$

Overfitting the observations → Finite Sample Undercoverage

Scaling the Residuals

Definition (Linear Estimator):

$$\hat{f}(\boldsymbol{x}^*) = \sum_{i=1}^{N} y_i w_i(\boldsymbol{x}^*)$$

$w_i(\boldsymbol{x}^*)$ depends only on $N, i, \boldsymbol{x}^*, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N$, $\sum_{i=1}^{N} w_i(\boldsymbol{x}^*) = 1$
In-sample matrix form:

$$\hat{f}(\boldsymbol{X}) = \boldsymbol{W}\boldsymbol{y}, \qquad \boldsymbol{W} = \begin{bmatrix} w_1(\boldsymbol{x}_1) & \cdots & w_N(\boldsymbol{x}_1) \\ \vdots & \ddots & \vdots \\ w_1(\boldsymbol{x}_N) & \cdots & w_N(\boldsymbol{x}_N) \end{bmatrix}$$

In-sample residuals:

$$\hat{\boldsymbol{\epsilon}} = (\boldsymbol{I} - \boldsymbol{W})\boldsymbol{y}$$
$$Cov(\hat{\boldsymbol{\epsilon}}) = (\boldsymbol{I} - \boldsymbol{W})\sigma^2 \boldsymbol{I}(\boldsymbol{I} - \boldsymbol{W})^T$$

Define $s_i = \left(\frac{1}{(1 - w_i(x_i))^2 + \sum_{k=1,k\neq i}^{N} w_k(x_i)^2}\right)^{\frac{1}{2}}$:

$$Var(s_i\widehat{\epsilon_i}) = \sigma^2$$

By the Lindeberg-Feller central limit theorem:

$$\frac{f(\boldsymbol{x}^*) - \hat{f}(\boldsymbol{x}^*)}{\|\boldsymbol{W}^*(\boldsymbol{x}^*)\|} \xrightarrow{d} N(0, \sigma^2), \qquad \frac{\hat{f}(\boldsymbol{x}^*) - \widehat{f_b}(\boldsymbol{x}^*)}{\|\boldsymbol{W}^*(\boldsymbol{x}^*)\|} \xrightarrow{d} N\left(0, Var(s_i\widehat{\epsilon_i})\right)$$

Denote a diagonal matrix $\boldsymbol{S}$ whose $(i,i)^{th}$ entry is s_i:

$$E\left(\frac{\boldsymbol{y}^T(\boldsymbol{I} - \boldsymbol{W})^T\boldsymbol{S}^T\boldsymbol{S}(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{y}}{N}\right)$$
$$= \sigma^2 + \frac{f(\boldsymbol{X})^T(\boldsymbol{I} - \boldsymbol{W})^T\boldsymbol{S}^T\boldsymbol{S}(\boldsymbol{I} - \boldsymbol{W})f(\boldsymbol{X})}{N}$$

Additional Bootstrap Alternative for Scaling

What if the regression estimator is black-box to the user?

Sample residuals:

$\widehat{\epsilon_i} = y_i - \hat{f}(\boldsymbol{x}_i)$, *empirically* $\widehat{F_N}$, *with sample variance* $\widehat{\sigma^2}$

Generate $y_{a,i} = \hat{f}(\boldsymbol{x}_i) + \epsilon_i,\ \epsilon_i \sim \widehat{F_N},\ a = 1,2,\dots,A$
$\widehat{f_a}(\boldsymbol{x}^*)$ is dependent on $(\boldsymbol{x}_i,\ y_{a,i})$ $i = 1,2,\dots,N$
Bootstrap sample residuals:

$$\widehat{\epsilon_{a,i}} = y_{a,i} - \widehat{f_a}(\boldsymbol{x}_i), \qquad \widehat{\sigma_i^2} = \frac{\sum_{a=1}^{A} \widehat{\epsilon_{a,i}}^2}{A}$$

Alternative scaling factors:

$$s_i = \left(\frac{\widehat{\sigma^2}}{\widehat{\sigma_i^2}}\right)^{\frac{1}{2}}$$

References (a selected list)

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-free predictive inference for regression. Journal of the American Statistical Association, 113(523):1094–1111.
Stine, R. A. (1985). Bootstrap prediction intervals for regression. Journal of the American Statistical Association, 80(392):1026–1031.
Tsybakov, A. B. (1986). Robust reconstruction of functions by the local-approximation method. Problemy Peredachi Informatsii, 22(2):69–84.
Tsybakov, A. B. (2009). Introduction to nonparametric estimation. Springer series in statistics. Springer, New York ;.
Van der Vaart, A. W. (2000). Asymptotic statistics, volume 3. Cambridge university press.