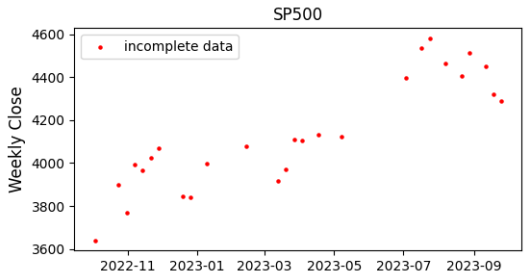
ACTU PS5841 Data Science in Finance and Insurance - Autumn 2023 (Y. Wang)
Assignment - 6
Assigned 10/19/23, Due 10/27/23

quadratic LOWESS with bootstrap CI

The file `SPX_W_incomplete.csv` contains weekly SP500 index data from the first full week of October 2022 to the last full week of September 2023. Each week is indicated by its Monday as well as its week number.

The file contains only 26 weeks instead of 52 weeks of data because half of the data are missing.



For your reference, the file `hint.html` contains python codes to generate a **full date grid**, containing every Monday during the period.

Let's code in python to use **the LOWESS model (with a weighted least squares fit of a quadratic regression in each local neighborhood)** to fit a curve through the weekly closes using the full date grid.

The local regression model is $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$, with the least squares solution $\hat{\beta} = (X^T W X)^{-1} X^T W y$, where X being the design matrix, and W being a diagonal matrix with the LOWESS weights in its diagonal.

- Please do NOT use functions from existing packages for the tricube function and LOWESS function. Please create your own code for these.
- Please use a neighborhood of K nearest neighbors for each target point, but include all furthest neighbors if their distances to the target point are tied. Thus some targets will have more than K neighbors.
- Please use a target grid of 52 weeks indexed by week numbers [0, 1, ..., 51]

Let's set $K = 15$
[a] Please let your code report the weeks in the target grid that have more than K neighbors by their Week Number, (Monday) Date, and Number of Neighbors.

Week Number	Date	Number of Neighbors
...	...	...

Please fit the data with the proposed LOWESS curve, and use the bootstrap approach to estimate the 95% confidence interval of the fitted curve.

- If a bootstrapped sample results in non-invertible $(X^T W X)$ for any grid point, you should abandon this "unusable" bootstrapped sample and move to the next one.
- You should estimate the confidence intervals based on at least 500 "usable" bootstrapped samples.

Please submit (*.ipynb and *.html) for your work

[*b*] Let your code report the fitted mean function and its confidence interval.

Week Number	Date	fit	CI low	CI high
...	...	...	...	...

[*c*] Please let your code visually compare your fit to that by `geom_smooth()`

```
geom_smooth(df, aes(x = 'WeekNumber', y = 'Close'),
method = 'loess', span = 0.60, se = True, method_args = {'degree':2,'level':0.95})
```