NYU COURANT

MATH-GA.2047-001 Data Science in Quantitative Finance
MATH-GA.2070-001 Data Science and Data-Driven Modeling

Homework 4

Profs. Ivailo Dimov and Petter Kolm
Due: October 6, 2020

Instruction

This homework is to be done individually. No collaboration and/or code sharing permitted.

Objective

In this assignment, you will:

- Perform linear regression analysis of (i) hedge fund returns and the Fama-French five-factor model and (ii) the housing dataset from class.

- Familiarize yourself with the standard data science workflow covered in class.

Methodology and Deliverables

1. Performance analysis of hedge fund returns using linear regression.

    (a) Download and import the following data into your Jypyter notebook:

    - Monthly data of the Fama-French 5 Factor model from Ken French's data library. The Fama-French 5 Factor model is described here.
    - Monthly returns from Yahoo Finance for the AQR Equity Market Neutral Fund Class I (Ticker: QMNIX). Make sure you get the series that is adjusted for both dividends and splits (i.e. the series marked "Adj Close**" on Yahoo Finance). Note that Yahoo Finance will provide you with prices. Hence you will have to compute returns yourself.

    Use the longest time period you can find where all data series are available.

    (b) Create scatter plots and compute summary statistics of all data series (mean, std, skew, kurtosis, max, min, correlation matrix). Discuss your conclusions and any noteworthy results after visualizing and exploring the data. (Extra credit: Perform tests for stationarity of all data series.)

(c) Regress QMNIX on the Fama-French factors and discuss the regression results, including t- and F-tests and R^2. Is the intercept statistically different from 0?

(d) Demonstrate graphically whether the classical linear regression assumptions are satisfied or not in (c).

(e) For students in the Data Science in Quantitative Finance class (extra credit for everyone else): Provide a financial/economic interpretation of your results from (c).

(f) Extra credit: Download and import the monthly data of the 10 Industry Portfolios model from Ken French's data library. The 10 Industry Portfolios model is described here. Then perform the analysis (c)-(e) using this data.

(g) Extra credit: Download and import the *daily* data of all the data series. Make sure you use the same time period as above. You will find them on Ken French's website and Yahoo Finance. Then perform the same analysis as above. Are the classical linear regression assumptions satisfied for the regressions with daily data? Does your analysis using daily data provide different results compared to the analysis based on monthly data? Why, or why not?

2. Linear regression models for the housing dataset.
   In this question you will be extending the Jupyter notebook

   `lec04_skl_ols_workflow_housing_data.ipynb`

   used during the last lecture.

   (a) Implement a `LinRegStatsmodels()` Predictor Class that uses the OLS statsmodels class and its `.fit()` and `.predict()` methods. Whenever you call `.fit()` to an instance `A` of your class, the statsmodels results should be recorded in an attribute `A.results`.

   (b) In the `full_pipeline_with_predictor` in the Jypyter notebook above, substitute your `LinRegStatsmodels()` class and then compute the summary statistics of your regression using both OLS and OLS with heteroskedastic standard errors.

   (c) Interpret your regressions results. What can you say about the models?