**DB Equity Quantitative Strategy – Data Exercise**

**Dataset:**

We have provided a csv file (db\_data.csv) that contains monthly factor data in the format of stock, date, fwd\_returns, f1, f2,….f20. We have obfuscated the data to preserve anonymity.

* The “fwd\_returns” field indicates the forward 1 month return for that stock on that date. When framing this as a prediction problem, this can be considered your “y” (or response) variable
* You may assume all dates are in continuous order (e.g. D002 comes after D001, but before D003)
* All data is “point-in-time”, meaning on date D002, you have access to all data on or before D002, but NOT after D002. This does not apply for “fwd\_returns” of course, since this is the forward 1 month return. If you lag it, it becomes the return for that particular month.
* Hint: factor F02 resembles a “sector” classification of each stock. Depending on your analysis, this may be useful to know.

**Instructions**

Given the factor data provided, please come up with a long/short multi-factor strategy of your choice that uses factors to select stocks. You may use R or python. (You don’t have to use all of the factors provided, you may use any number of factors you wish.)

You may want to split up your work into different stages:

* Data cleaning / wrangling / dealing with outliers, missing data, etc
* Exploratory data analysis
* Model development / predictive modelling – you may use any modelling techniques you wish
* Backtesting / portfolio construction / evaluation of trading strategies [NOTE: when backtesting your strategy, be sure to lag the “fwd\_returns” so that you get contemporaneous returns]

Time limit – please email back all code and your write-up within **3 hours.**

Be mindful of:

* Lookahead bias - the data provided is point-in-time, so you need to be careful not to use future factor / returns data. For example, in your strategy, you may not use factor data from D003 to trade on D002. We will check your code for instances of look-ahead bias.
* Overfitting - though we did not split up the data into test / training set, we encourage you to do so

Submission:

* Please provide a write up of your analysis and findings (1-3 pages)
* Please document all code to make it easy for us to follow
* Please submit all code used for your analysis into a zip file (while we may not run your code, you should assume that we will)
* You may use ipython notebook or Rmarkdown to present your results together with your code
* Please do not include any compiled code as these will likely trigger spam filters

Evaluation

* We will take everything into account in our evaluation
  + Coding skills (clarity, design, well-documented, understandable)
  + Data analysis
  + Thought process
  + Writing skills
  + Knowledge and ability to implement modelling techniques
* We want to make this as open-ended as possible. There are no right or wrong answers. We just want to see what you come up with. Whether or not you come up with a high sharpe strategy is not as important as your thought process.
* If you run out of time and didn’t get a chance to complete what you intended, please write down things you would do if given more time

Questions

* If you have any questions, please use your best judgment and write down your assumptions
* Feel free to reach out to us if you have urgent questions. We’ll try our best to answer them as quickly as we can.

Good luck!