Yuan Chen
ECE 532
UW-Madison

Predict the Stock Market Return Using Micro-Economic Indicators

Yuan Chen

Problem:

Equities, or stocks, are the most popular asset class that attracts numerous individual and wholesale investors. An enduring topic is how to accurately predict the movement of the stock market. Although the market fluctuation seems random most of the time, it is highly correlated to macroeconomic conditions. Therefore, most macro trading strategies are based on the various combination of some indicators.

Here I propose to use three macroeconomic indicators base on interest rates, key factors that have significant impact on the economy - default spread, term structure spread, and the risk-free rate, to predict the stock market return.

Default spread is defined as the yield gap between AAA and BBB grade bonds. AAA bonds are issued by the companies with great outlook, so it has the lowest risk thus lowest yield. And BBB bonds are those with a higher risk of default, yet still, have investment value. When the economy is at stake, investors seek a higher and higher return to pay off the increasing risk.

The term structure spread is the yield curve of bonds, which demonstrate the interest rates of bonds with the same quality at different maturities. The slope of the curve is a crucial pointer of the current state of the economy. An upward sloping means the long-term yields are higher than the short-term ones point to an expanding economy and vice versa. A flat cure means the economy is stalling or the market is unsure of the future.

A risk-free rate is considered as the baseline of investment since they will only invest if the return is greater than the risk-free rate. The three-month U.S. Treasury bill is usually used for the calculation for the fact that it has the lowest default risk.

Data set:

In this project, 20 years of data with a daily report from 2000 to 2019 will be download from (Kenneth French Data Library: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html and Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/). There are ~250 trading days each year, so about 5000 data points with the three indicators will be stored in a 5000X3 matrix. The historical return of the stock market of every month following each data point will be stored in a 5000X1 vector, labeled as +1 if the return is positive, and -1 if the return is negative. This vector will be used to compare with the approximation.

Algorithm:

Three training algorithms I will use in this project: Ridge Regression, LASSO, and Neural Network.

Yuan Chen
ECE 532
UW-Madison

Ridge Regression: the ridge parameter λ is the main variable in this method, I can start with a few random value and tune it until a local/global minimum is found.

LASSO: although LASSO differs from Ridge Regression by imposing an L1 penalty instead of adding an L2 penalty. Tuning λ is still the main task to train the algorithm.

Neural Network: there are many configuration options in the Neural Network to fine-tune the algorithm. To accomplish the project in a reasonable time frame, I will fix the number of layers and nodes, and activation function, mostly change the learning rate, momentum, iterations to compare the error of the result.

The algorithms will be evaluated by K-fold cross-validation. The data set will be partitioned into 20 subsets with one year's worth of data in each. The training and validation process will be executed 20 times, in each time, one subset will be used for validation and the rest are used for training. The average error will be used to compare the algorithms.

If time allows, repeated random sub-sampling cross-validation may also be used to validate the training algorithm. Similar to the K-fold cross-validation, but the difference is each time the training and testing subset is chosen randomly.

Github:

https://github.com/yuanchenGH/ML_MarketPrediction

Timeline:

Oct 22: submit the proposal for the final project.

Oct 22 – Oct 29: collect and clean the data set, make a framework for the learning algorithm.

Oct 30 – Nov 16: write algorithms to train the data set, obtain preliminary results for at least two algorithms. Analyze the result and look for possible improvements. Revisit the learning plan and make changes if necessary.

Nov 17: write and submit the first update report.

Nov 18 – Nov 30: finish the writing of all three algorithms, in the meantime, vigorously test different models and parameters.

Dec 1: write and submit the second update report.

Dec 2 – Dec 11: wrap up the project, including clean up code and make sure the final results are free of errors if possible. Write the final report.

Dec 12: submit the final report.

Dec 13 - Dec 17: review two projects from other students and write review reports.