Long-Short Strategy Using Bank Analysts Recommendations

CS 229 - Project

Thomas AYOUL, Alexandre ELKRIEF & Nicolas MEILLER

Abstract

In this project, we aim at trying to predict stock prices of the Eurostoxx50 index using machine learning techniques applied to bank analyst recommendations. Based on these predictions, we create a long-short trading strategy. We then back-test this strategy and use different validation tools to improve the model. We investigate two types of logistic regression models which achieve a 43% and 33% error rate respectively. We then build strategies using sliding training set based on SVM, Random Forests. Finally we add trend following and mean reversion features.

Data Acquisition

- Pulling bank analysts ratings and target prices for each stock composing the Eurostoxx50.
- Data from January 2000 to June 2015.
- Reformattting and concatenating the time series into data-frames

Model 1 : Naive Logistic Regression

- Long-short strategy requires buy and sell signals.
- Model : Logistic regression.
- Assumption : independent increments and same behavior.
- Stocks modeled as a single response vector.
- Training size : 75%.
- Features and Label : $Y = \begin{cases} 1 & \text{if } \text{Eurostoxx}\text{Return}_{t+1,n-1} > 0 \\ 0 & \text{otherwise} \end{cases}$

Model 2 : Greedy Logistic Regression

- Features and Label : $X = [X_1, X_2, \ldots, X_{50}]$
- Constraints : $X_i = \begin{cases} \text{StockReturn}_i, \text{AnalystTargetReturn}_i, \text{RatingReturn}_i & \text{if } \text{Eurostoxx}\text{Return}_{t+1,n-1} > 0 \\ \text{AverageRating}_i, \text{EurostoxxReturn}_i & \text{otherwise} \end{cases}$

Model 3 : Sliding logistic regression

- Features and Label : $X = [X_1, X_2, \ldots, X_{50}, \text{Eurostoxx}_n]$ with $X_i = [\text{StockReturn}_i, \text{AnalystTargetReturn}_i, \text{RatingReturn}_i]$
- New parameters : size of sliding window, length of moving averages.

Model 4 : Comparison of classifiers

- New classifiers : Random Forests and SVM

Conclusion:

- Error rate : 43%
- Could work great (Nokia) but could lose a lot (Carrefour)
- Too strong assumption : no industry specific components and independent stocks.
- Training set inadequate : using information from the year 2003 to predict 2014.

Performance

- Error : 33%
- Sufficiently good to have positive excess return over the Eurostoxx.
- $\alpha = 10\%$ over last year.

Remark:

- The best window size corresponds to the analyst recommendations time scale.

Comparison Long-short vs Long-only strategy

Conclusions

- Applying machine learning techniques to the stock market seems to be performing well. Suggesting that further study in the domain, in particular on different and maybe less liquid data sets could lead to the finding of good strategies.
- The inclusion of bank analyst recommendations which are based on fundamental economic valuations allows for the strategies to incorporate a different and more fundamental approach than exploiting only the time series data. Again, further study on the inclusion of fundamental parameters in algorithmic trading strategies could lead to interesting results.