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 those 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.

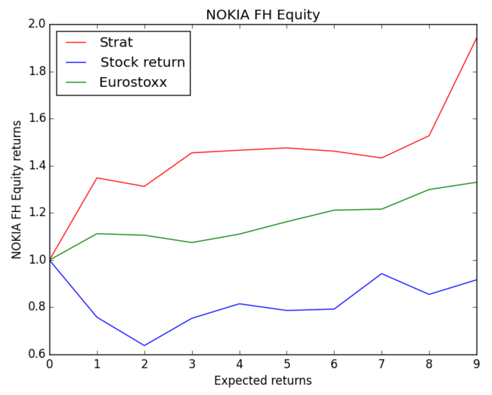
Data Acquisition

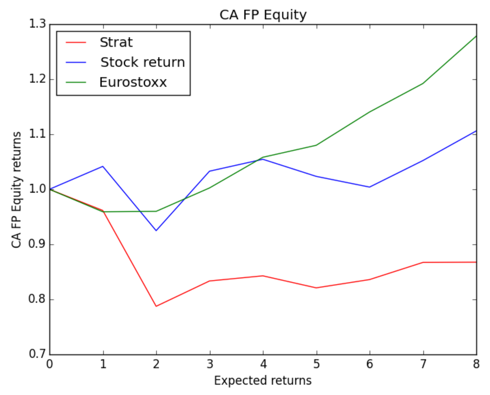
- Pulling bank analysts **ratings** and **target prices** for each stock composing the Eurostoxx50.
- Data from January 2000 to June 2015.
- Reformatting and concatenating the time series into data-frames

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 : $Y = \mathbf{1}_{\{StockReturn_{t+n_{day}} > 0\}}$

$$\mathbf{X} = [\, StockReturn_t,\ AnalystTargetReturn_t,\ RatingReturn_t,\ AverageRating_t,\ EurostoxxReturn_t \,]$$





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.

Model 2 : Greedy Logistic Regression

- New point of view : **predict Eurostoxx**
- More features and **correlation** taken into account.
- Training size : 90%.
- Features and Label :

$$\mathbf{X} = [X_1,\ X_2,\ \ldots,\ X_{50}]$$

with $X_i = [\, StockReturn_t^i,\ AnalystTargetReturn_t^i,\ RatingReturn_t^i \,]$

$$Y = \begin{cases} 1 \text{ if } EurostoxxReturn_{t+n_{day}} > 0 \\ 0 \text{ otherwise} \end{cases}$$

Performance

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

Remark

- **Dependent** on **training size**

Model 3 : Sliding logistic regression

- Addressing model 2 flaws: including **sliding training set window**.
- **Including moving average** to incorporate **trend following** and **mean reversion** features
- Features and Label :

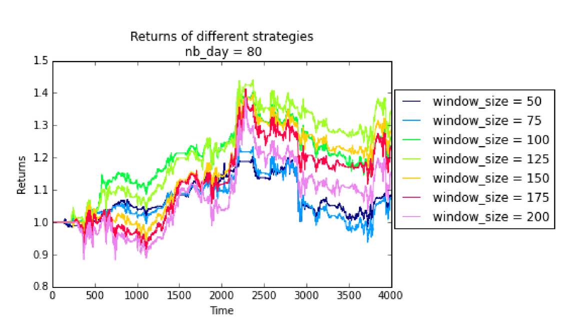
$$\mathbf{X} = [X_1,\ X_2,\ \ldots,\ X_{50}, Eurostoxx_{ma}]$$

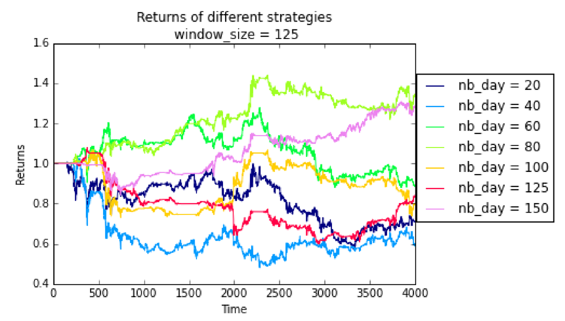
with

$X_i = [\, StockReturn_t^i,\ AnalystTargetReturn_t^i,\ RatingReturn_t^i \,]$

- New parameters : size of sliding window, length of moving averages.

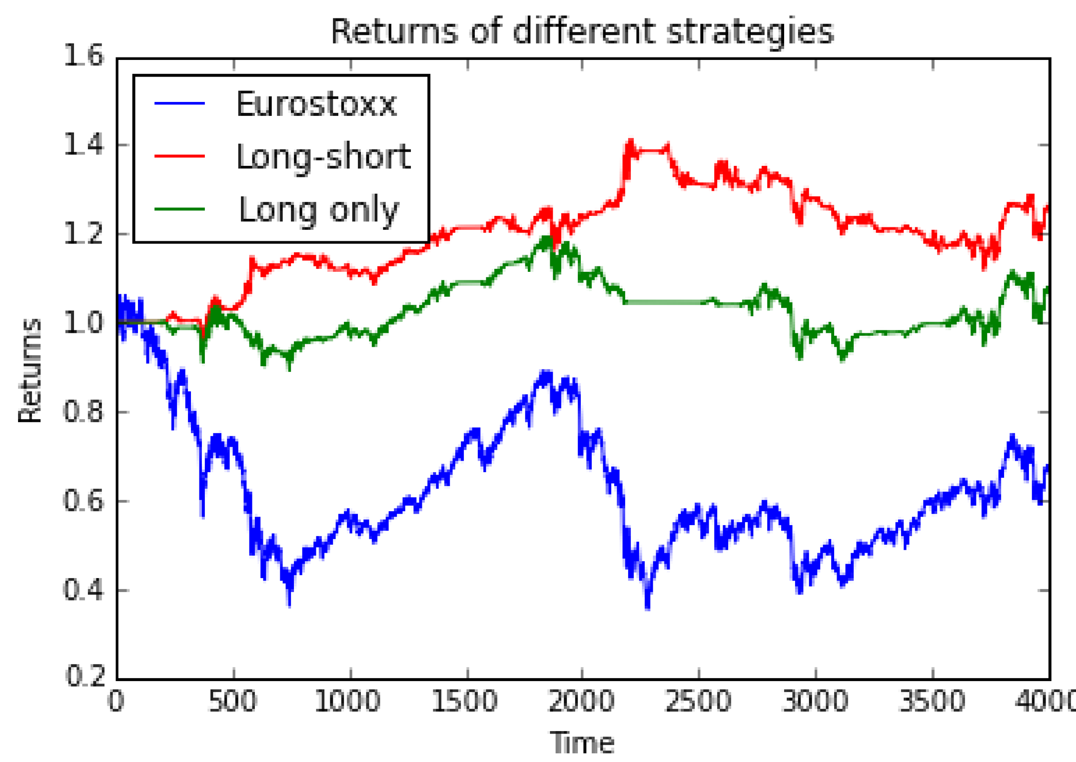
Optimization of parameters





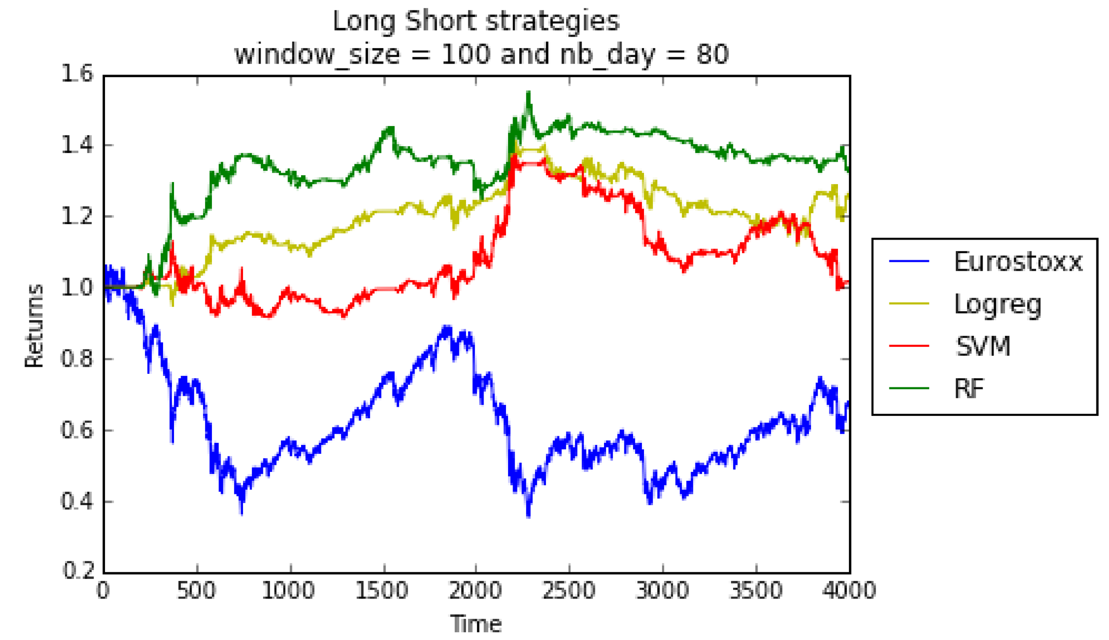
Remark : The best window size corresponds to the **analyst recommendations time scale**.

Comparison Long-short vs Long-only strategy



Model 4 : Comparison of classifiers

- New classifiers : **Random Forests** and **SVM**



Results

- Long-short vs long only strategies: **long-only performs less well** than the long-short strategy for all classifier which validates the ability of the strategies to **profit even in difficult market situations**.
- Different classifiers in different scenarios: SVM works better to predict brutal crashes such as the 2008 crisis, performs less well than the logistic regression and random forests over longer and less volatile periods. Indeed the latter deliver returns more consistently.
- Volatility of strategies:

Classifiers	Annualized volatility
Logistic regression	9.0 %
SVM	10.9%
Random Forest	8.50 %
Eurostoxx	14.5%

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 analysts 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**.