What factors predict S&P500?

J. Cavegn, A. Falk, J. Raschke

UZH, Department of Banking and Finance

2023

University of Zurich

Table of Content

University of Zurich™

List of abbreviations

- DTC decision tree classifier
- RFC random forest classifier
- SVM support vector machine

University of
Zurich™

Introduction

- Prediction of S&P 500 returns with an emphasis on currency data
- Examine whether fluctuations in exchange rates can serve as a real-time signal for future S&P 500 returns
- Three supervise machine learning models used to capture possible complex interplay between FX rates and the equity markets:
    - Support Vector Machine
    - Decision Tree Classifier
    - Random Forest Classifier
- Detailed information about the implementation (functions and code) can be found in the full report

University of Zurich™

Support Vector Machine

- Supervised machine learning model for classification and regression tasks
- Idea: Find the hyperplane that best separates the data into distinct classes while maximizing the margin between them
- SVM's ability to handle high-dimensional data and capture non-linear relationships makes it a valuable tool for investigating the intricate relationships between FX rates and equity markets

University of Zurich™

Decision Tree Classifier

- Decision trees are hierarchical tree-like structures that recursively split the dataset based on feature conditions
- Idea: Splits are determined by selecting the features that maximize information gain or minimize impurity at each node
- Decision trees are intuitive models that can capture complex decision boundaries and are easily interpretable markets
- Parameters: Maximum depth, which controls the complexity of the model

University of Zurich™

Random Forest Classifier

- Random Forest Classifier combines several decision trees and combines their predictions
- Idea: Each tree is trained on a random subset of the data, and the final prediction is obtained through a voting mechanism
- Random Forest enhances model robustness and generalization by mitigating overfitting. By aggregating the outputs of individual trees, we aim to improve the overall accuracy and reliability of our predictive model.
- Parameters:
    - Trees: Number of decision trees in the "forest"
    - Maximum depth: The maximum depth of one decision tree in the forest
    - Leaves: Number of terminal nodes in one tree

University of
Zurich™

Initial Results

- The idea of all three models is to predict an up or down-move of the S&P 500, in which case they can take a long or short position to profit from the anticipated move
- Initially, the three models were trained with all available currencies as predictor variables, a lag of one day, 0.8 training/test set split, allowing both long and short positions. All other parameters were set to their default values as specified in the report.
- Prediction accuracies are displayed in the table below:

Classifiers	In Sample	Out-of-Sample
SVC	0.581024	0.510906
DTC	0.624265	0.477349
RFC	0.564442	0.510067

Table: Performance of Classifiers

University of Zurich™

Initial Results

- As can be gathered from the table in the previous slide, out of sample performance is not better than flipping a coin.
- Additionally, as seen in the figure below (GIF) below, RFC and DTC backtesting performances are heavily seed-dependent

Figure: Cumulative returns for different seeds.

University of Zurich™

Simulation

We simulate 250 times for the decision tree and random forest model. Unfortunately due to our simplistic architecture in our SVM, it only produces one possible path. From our simulated results the key findings are:

- There is lots of variability in performance
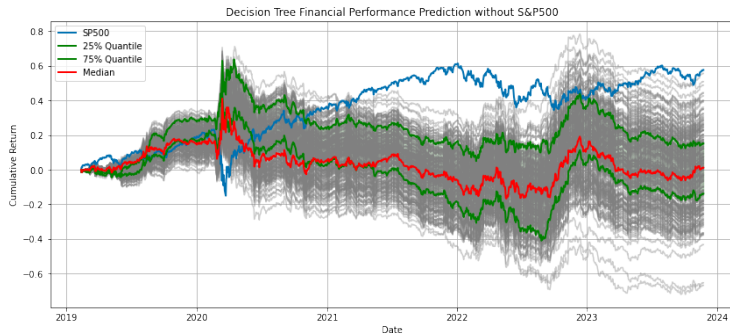- Random Forest significantly outperforms the decision tree

University of
Zurich™

Decision Tree Simulation Results

Figure: Decision Tree Simulation

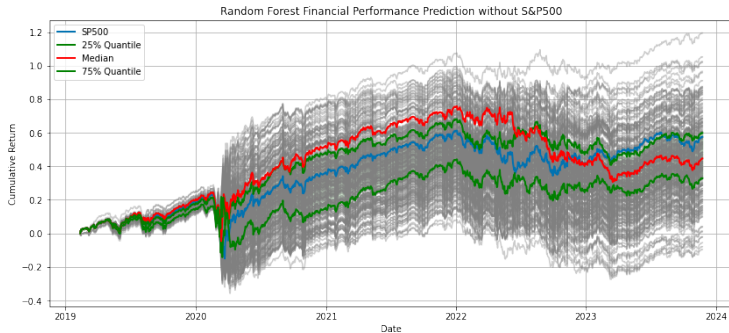# Random Forest Simulation Results

Figure: Decision Tree Simulation

Finding the best predictor (SVM)

- Here we analyze the best currency combinations (as predictor variables) ranked by backtesting performance for the SVM model
- With 10 different currencies, there are 1023 combinations of currencies. The highest performing combination with a return of 114% is EUR, JPY, CAD, SEK and SGD
- The reason for the high returns is due to being short the S&P 500 during the 2020 COVID-19 stock market crash

University of
Zurich™

Finding the best predictor (SVM)

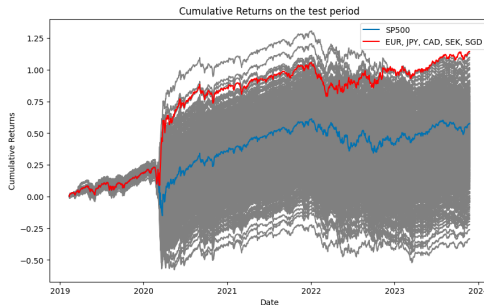- The following plot shows the 1023 predictor variable combinations with the aforementioned best one marked in red:

Figure: Cumulative returns of all different currency combinations used as explanatory variables.

Finding the best predictor (SVM)

- The figure below shows the distribution of the cumulative returns of backtests of all currency combinations:
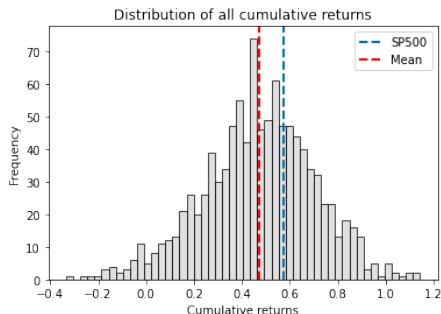

Figure: Cumulative returns of all different currency combinations used as explanatory variables.

Appendix A - Playground tool

- In the git repository, an interactive playground tool is included
- The tool is run in a jupyter notebook and allows the user to modify all parameters and the underlying dataset of the analysis in real time
- To set up the tool, follow the docker tutorial in the git repository

University of Zurich™