DEVELOPING A ZERO-INVESTMENT TRADING STRATEGY

Anjali Chauhan, Anjali Tiwari, Catherine Li, Jeremy Lee, Nish Patel

DIFFERENTIAL CAPITAL

Project Overview

OBJECTIVE

Identify the most promising portfolio of stocks to buy and short on a daily basis

Data: Daily stock prices from the S&P 500 from 1990-2021

↓

Generate risk-free returns (in theory)

TERMINOLOGY

Long: Buying a stock

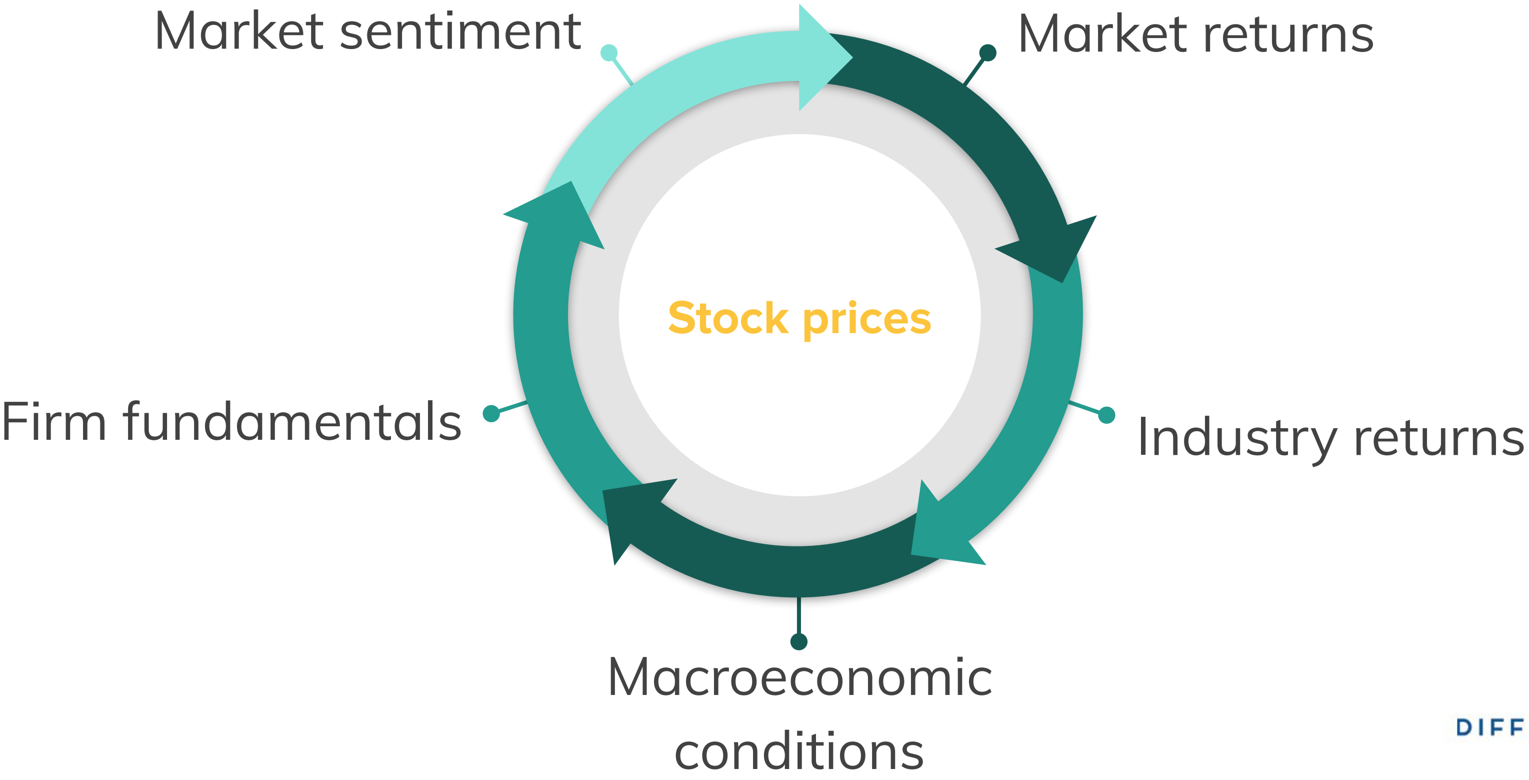
- Bet that price will increase
- Spend money now

Short: Selling a stock before owning it

- Bet that price will decrease
- Receive money now

Zero-investment: Using proceeds from shorting the worst-performing stocks to buy the best-performing stocks *net investment = 0*

What drives stock prices?



The million-dollar question

Can we predict the question?

Can we predict the market?

Efficient Market Hypothesis: Stock prices already reflect all available information

- **Implication:** Price changes are random and impossible to predict consistently
- **Fundamental assumption:** Markets are efficient

Reality: Markets are not efficient

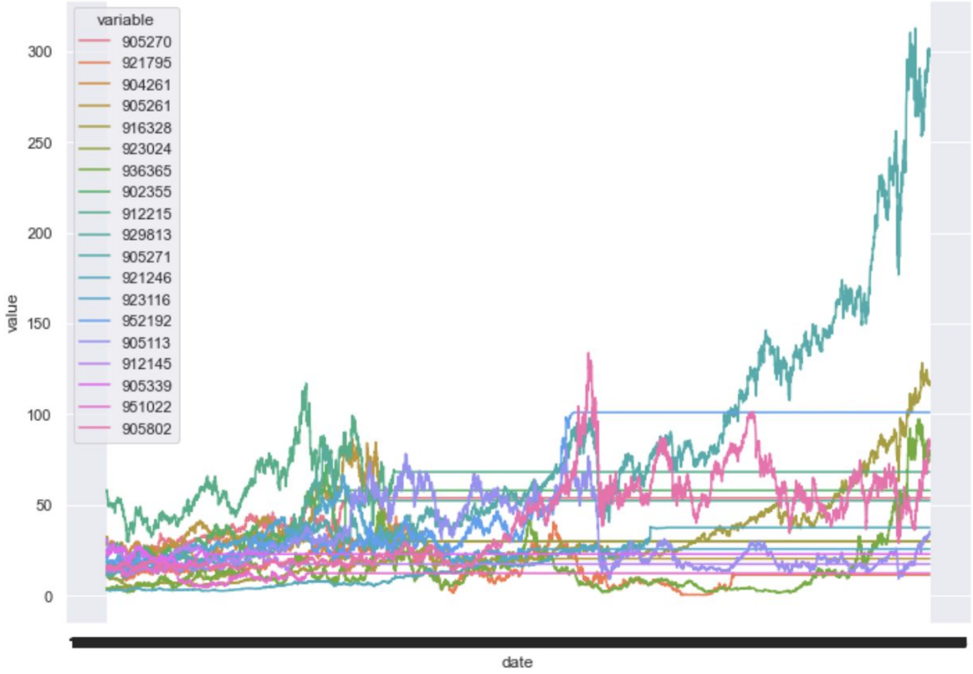
- Prices do not immediately adjust to new information
- There exist correlations between successive price movements while price adjusts

Data

7,914 time periods
1,214 stocks

Missing Data:

- When stocks were not listed on the S&P
- Could not drop or replace



DIFFERENTIAL CAPITAL

Data Preparation

Converted price information into stock returns

Question: What are our predictors and target variables?

For a given stock **X**:

- **Inputs:** Each stock's return at time **t**
- **Target:** Stock **X**'s return at time **t+1**

DIFFERENTIAL CAPITAL

Approach

for each time period, **t**: (3,627)

find stocks in the investable universe at time **t** (stocks in the S&P that have prices recorded for the last **200** days) (500-800)

for each stock, **s** in the investable universe:

X = lagged return of all stocks in universe

y = one-day-ahead returns of **s**

fit model and use to predict return of **s** at **t+1**

find tomorrow's best 5 stocks and worst 5 stocks (according to predictions)

portfolio return = average *actual* return of best 5 stocks - average *actual* return of worst 5 stocks

Feature Engineering

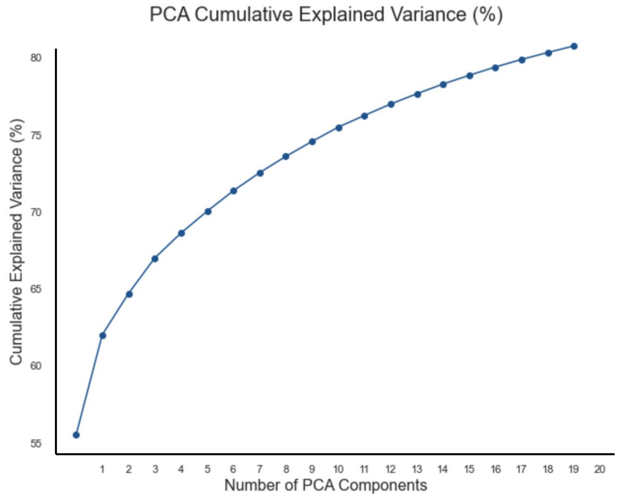
Working with 500+ features

- Not all past stock returns play an equal role in predicting stock **X**'s return tomorrow
- If stock **Y** and stock **Z** are correlated, taking both into account is redundant
- Models take a long time to train

Dimensionality Reduction

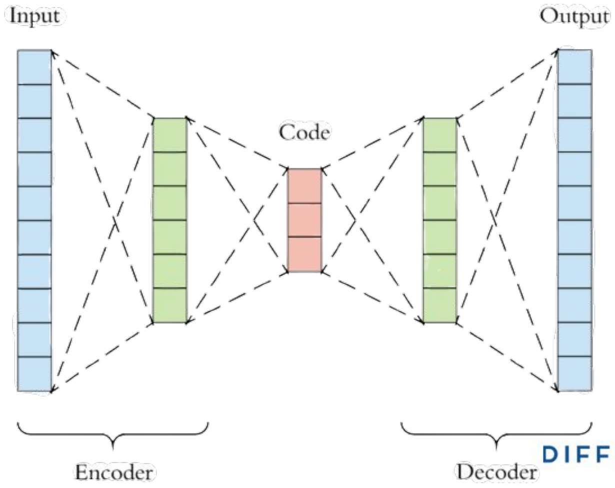
PRINCIPAL COMPONENT ANALYSIS

Creates new uncorrelated components (linear combinations of original variables) that explain the most variation in future returns



AUTOENCODERS

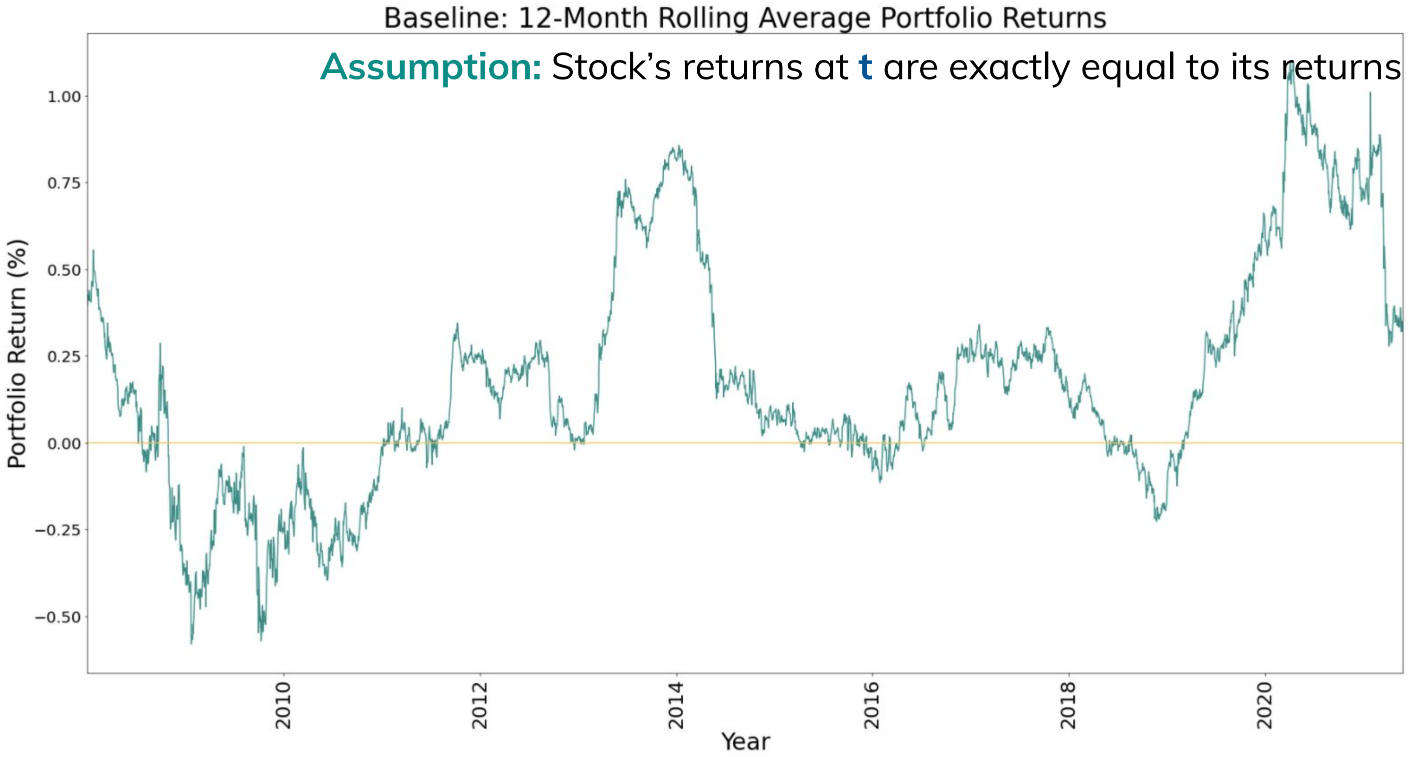
Artificial neural networks that learn a compressed representation of input data



DIFFERENTIAL CAPITAL

Regression Models

Baseline Model



Assumption: Stock's returns at **t** are exactly equal to its returns at **t-1**

STATS

Mean: .19%

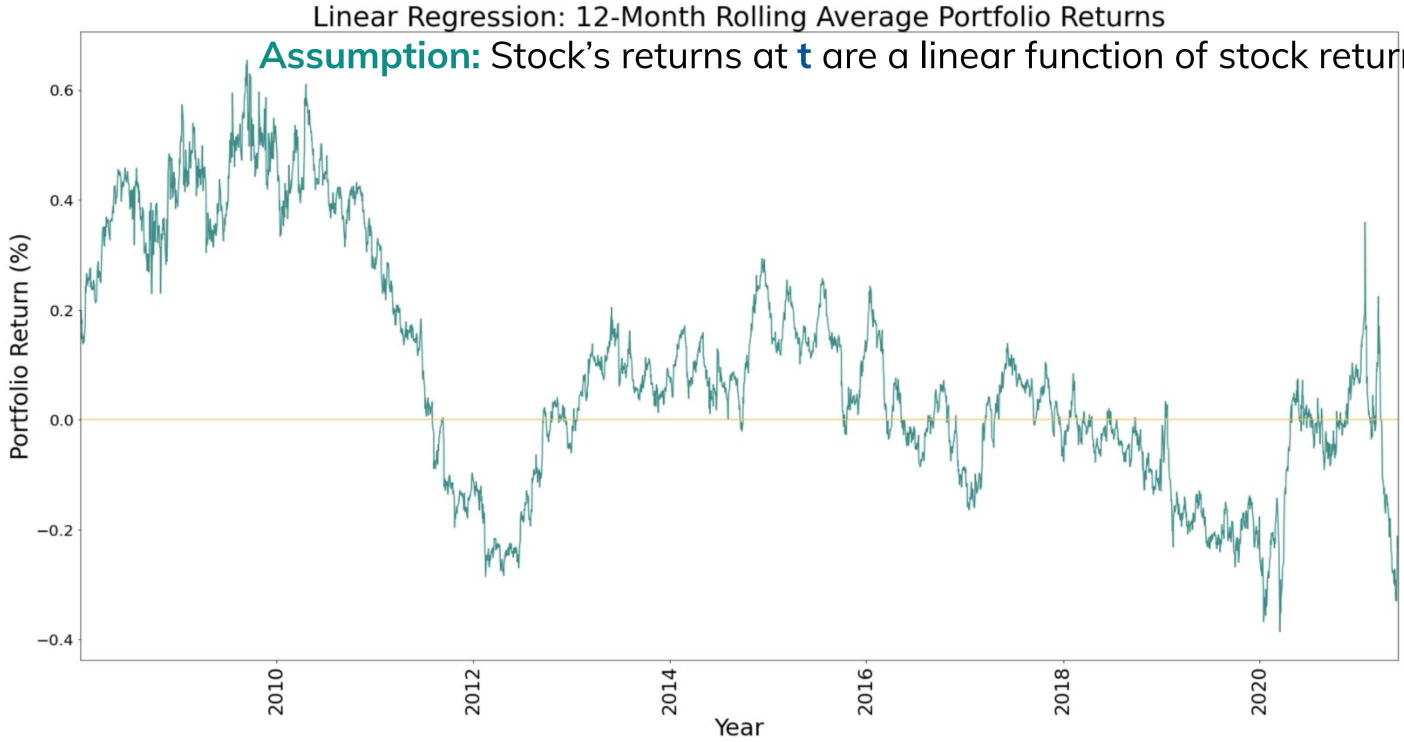
SE: .069%

95% CI

[-.017%, .397%]

DIFFERENTIAL CAPITAL

Linear Regression



Assumption: Stock's returns at **t** are a linear function of stock returns' at **t-1**

STATS

Mean: .08%

SE: .059%

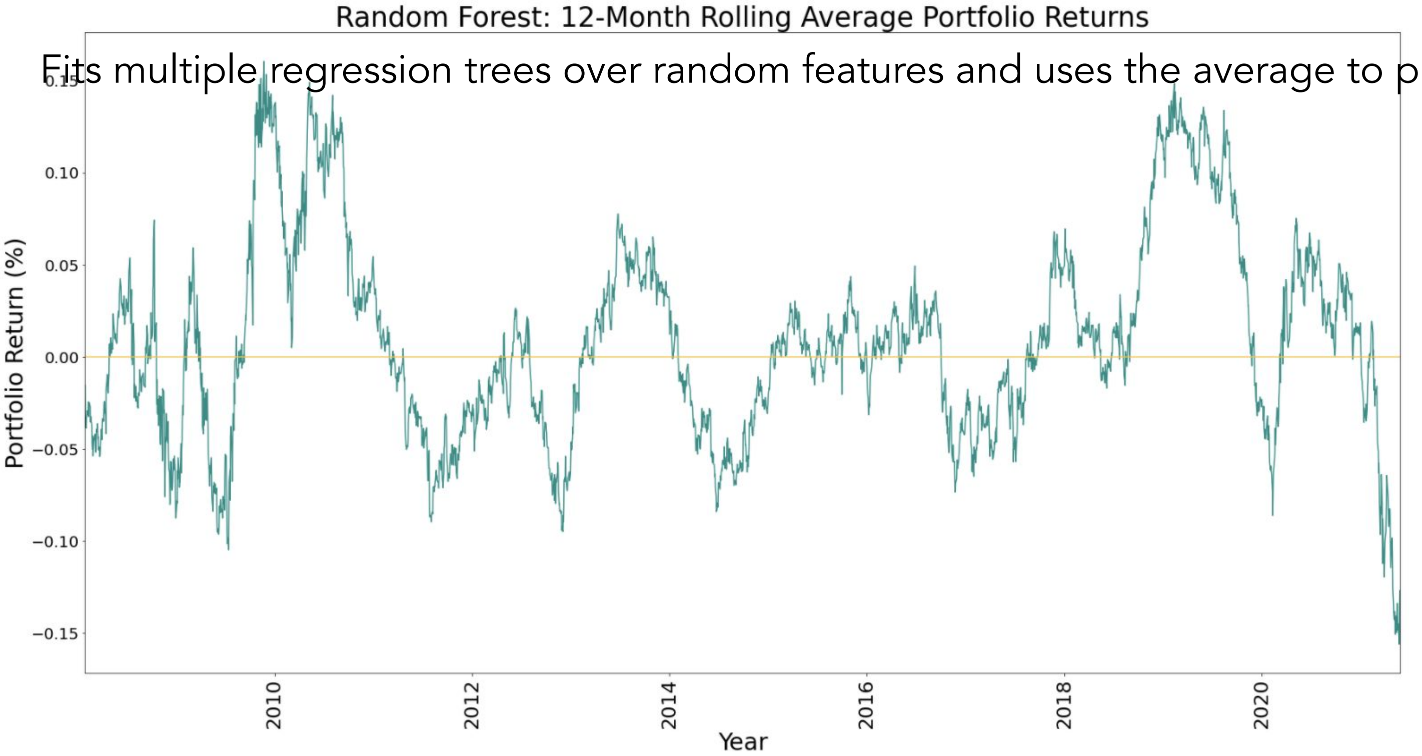
95% CI

[-.097%, .257%]

DIFFERENTIAL CAPITAL

Random Forest

Fits multiple regression trees over random features and uses the average to predict the returns



STATS

Mean: .0046%

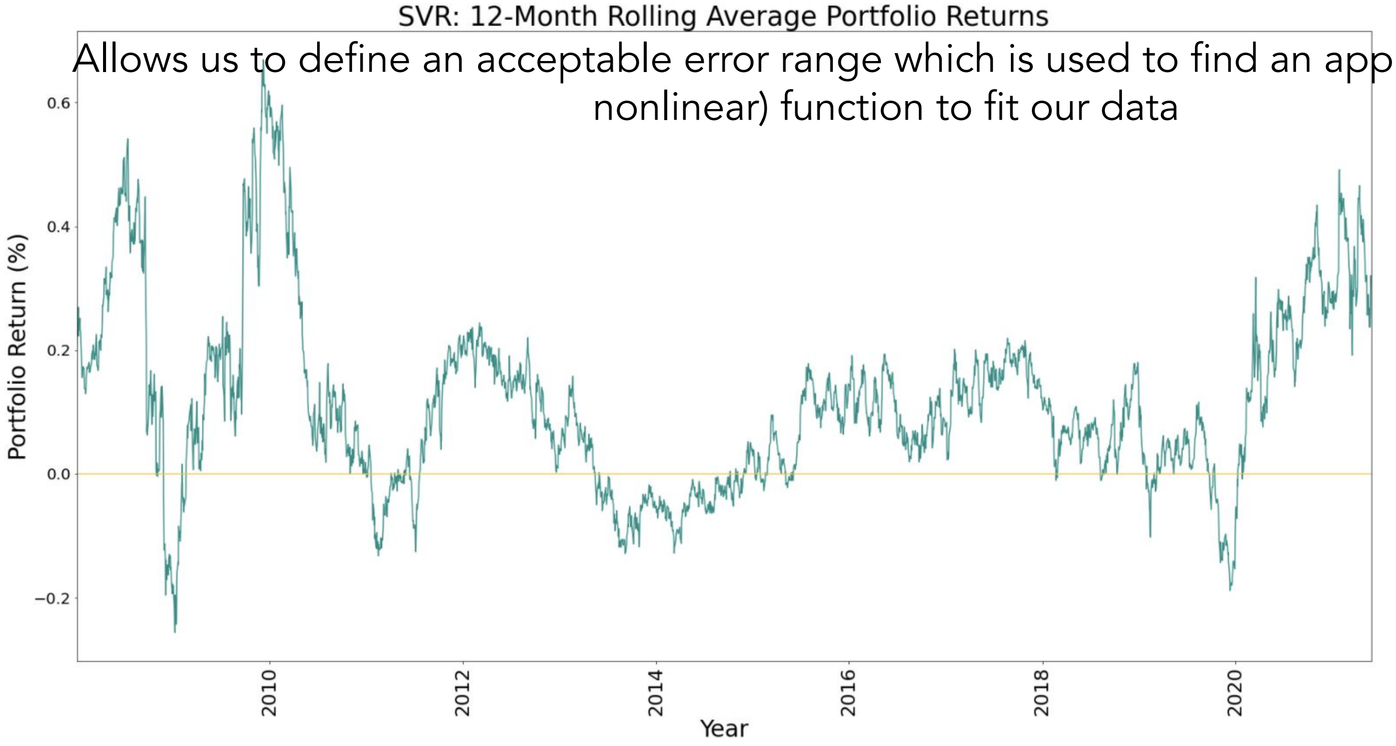
SE: .001%

95% CI

[.0016%, .0076%]

DIFFERENTIAL CAPITAL

Support Vector Regression



Allows us to define an acceptable error range which is used to find an appropriate (possibly nonlinear) function to fit our data

STATS

Mean: .12%

SE: .053%

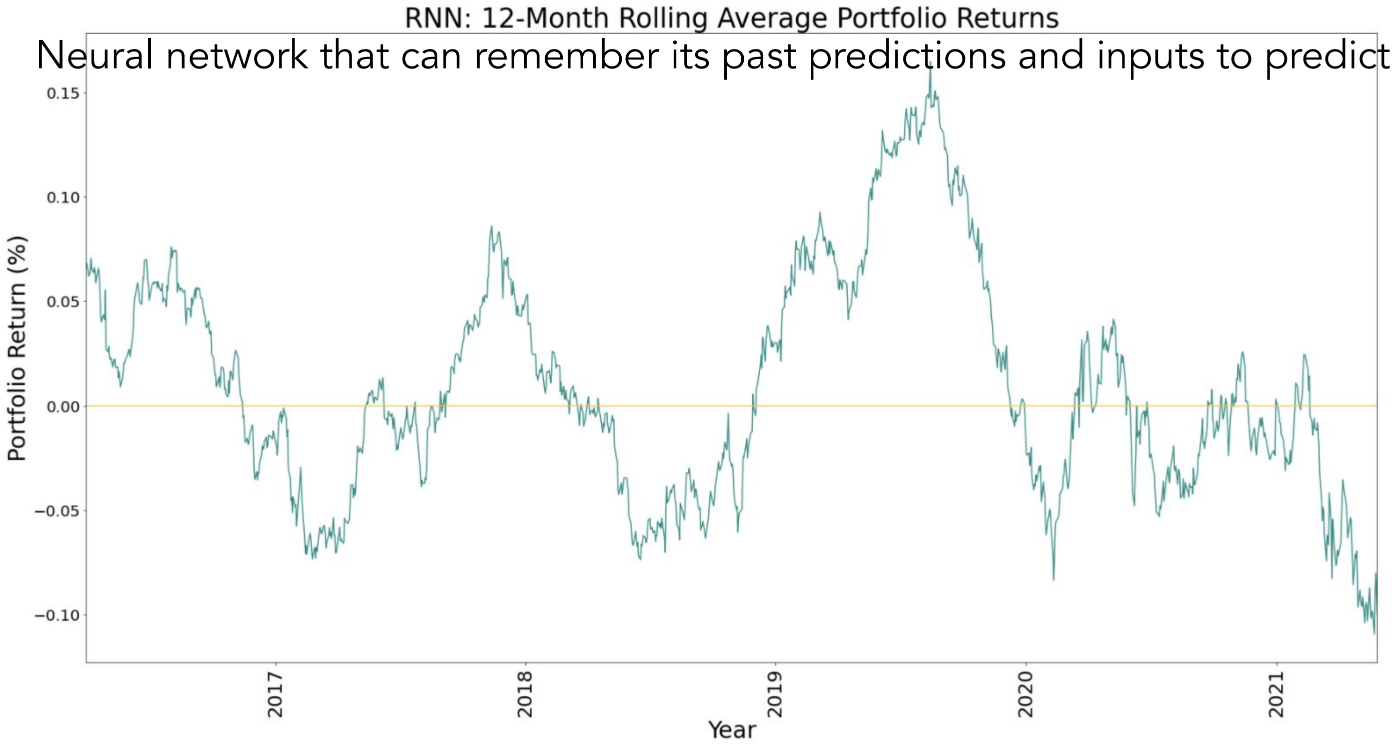
95% CI

[-.039%, .279%]

DIFFERENTIAL CAPITAL

Simple Recurrent Neural Network

Neural network that can remember its past predictions and inputs to predict future values



STATS

Mean: .0015%

SE: .028%

95% CI

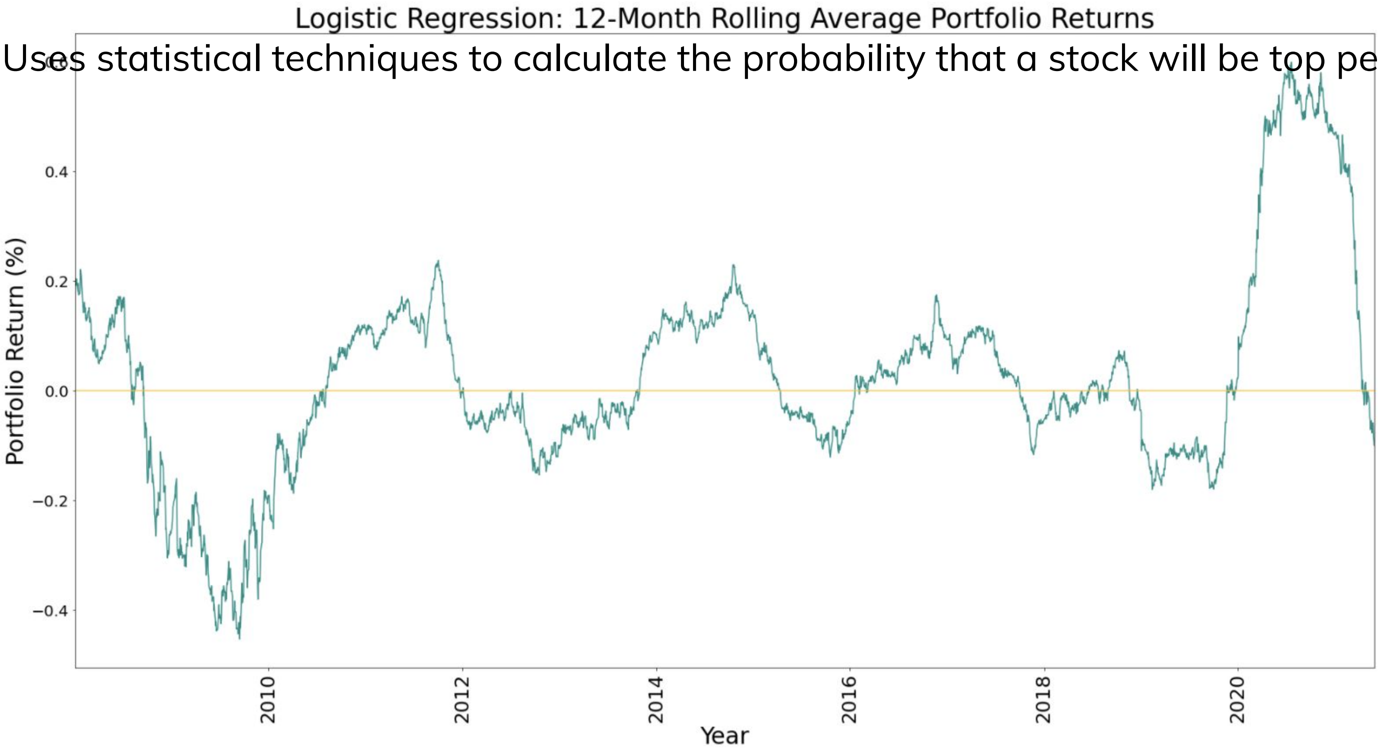
[-.0825%, .0855%]

DIFFERENTIAL CAPITAL

Classification Models

Logistic Regression

Uses statistical techniques to calculate the probability that a stock will be top performing stock at **t+1**



STATS

Mean: .02%

SE: .03%

95% CI

[-.07%, .11%]

DIFFERENTIAL CAPITAL

Support Vector Machines

Uses geometric techniques to separate worst-performing and best-performing stocks as distinctly as possible to predict the probability that a stock will be top performing stock at **t+1**

STATS

Mean: .04%

SE: .023%

95% CI

[-.065%, .073%]

DIFFERENTIAL CAPITAL

Conclusion

MODEL EVALUATION

- Models did not beat the baseline model
- Models cannot accurately form a portfolio that consistently generates positive returns
- Efficient Market Hypothesis could be stronger for the US market

FURTHER WORK

- Finetune autoencoder and investigate its effect on portfolio returns
- Tune hyperparameters for each stock model on the cloud to reduce the training time
- Use more frequent data (hourly) to see if prices are slower to adjust to new information

FINALLY,

- Thank you, **Miguel!**
- Thank you, **Enock!**

DIFFERENTIAL CAPITAL

Questions?