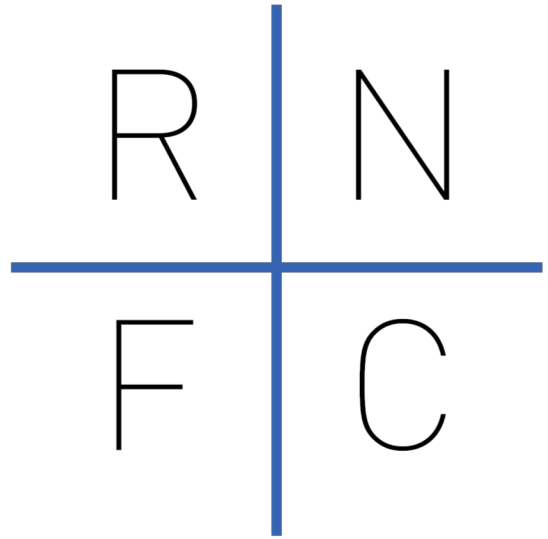
Factor Investing Using Deep Learning

Steven E. Thornton – Data Scientist, RN Financial Corporation

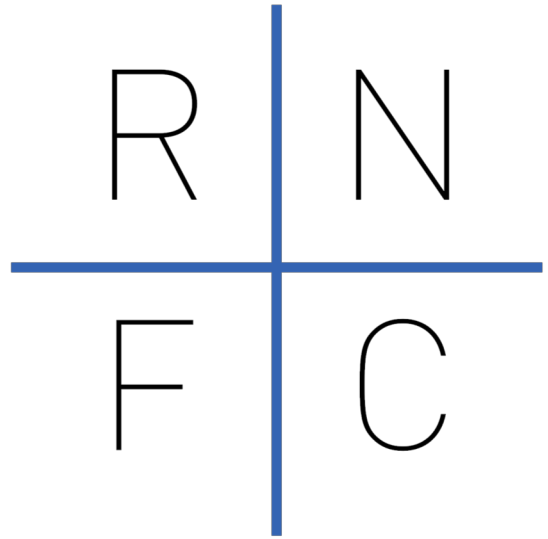
Rafael Nicolas Fermin Cota – CEO, RN Financial Corporation



RN Financial Research Centre is a financial research company focused on developing high-performance software for generating clean data. We use data science for quantitative analysis on equities by leveraging fundamental company data.

Part One - Data

"Data is the new Oil"



Data quality

Quality of data is paramount

Free of biases

Data needs to be clean

Right format

Homogenized

Trustworthy source

Implementable

Up-to-date

Data we use

Descriptive

Pricing

Fundamentals

Corporate Action

Research process

1. Obtain data
2. Clean data
3. Linking with other datasets
4. Generating aggregations or new features
5. Storing in a database

Issues we overcame through our experience

1. Extraction is slow
2. Messy
3. Inflexible
4. Difficult to compare among analysts
5. Cleaning is slow

Domain expertise

We've experienced the headaches of data cleaning

We're not just a data vendor

Constantly refining our process

We use our data

Software and hardware

Software

- Retrieve data from vendors
- Distributed analytical database
- High performance C++ for data cleaning
- Easy to add new factors
- Scalable

Hardware

- High-performance hardware managed by experts
- Storage systems

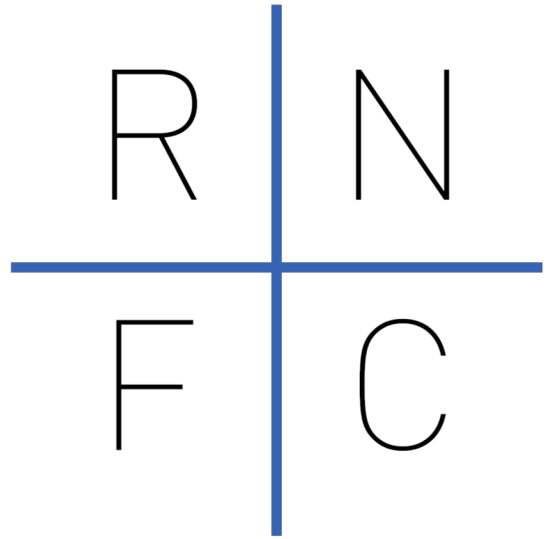
Skill management

Developer

Financial Expert

Part Two- Case Study

Applying Deep Learning for Fundamental-Based Signal Generation



Objective

Build a model that forecasts if stock ABC will outperform stock XYZ over the next year.

Optimized using GPUs

1. Model Training

2. Portfolio Construction

Data

1,233 Factors

- Momentum
- Value
- Growth
- Technical
- Sentiment

Investable Universe

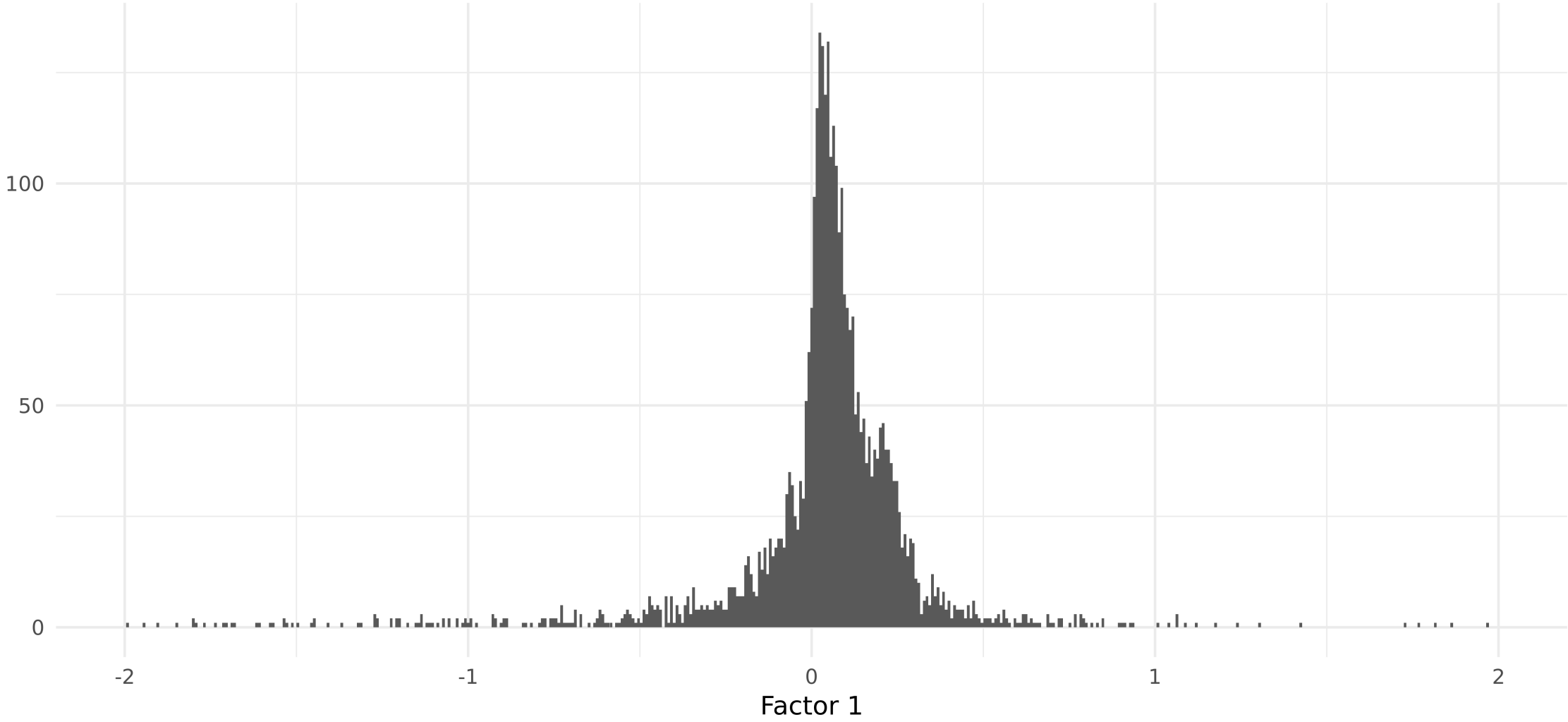
- Market cap of at least $1B USD
- Price of at least $5.00
- Sector Filters
- Corporate action filters

Investable universe

Model overview

1. Rank factor values on each day
2. Generate all pairs of stocks on each day
3. Predict if stock ABC will outperform stock XYZ over the next year
4. Generate a probability matrix
5. Compute the probability of each stock outperforming/underperforming all other stocks
6. Long the top 50 stocks each day, short the bottom 50
7. Apply a portfolio construction algorithm each day to determine stock weights
8. Layer portfolios each day

Step 1: Rank factor values on each day

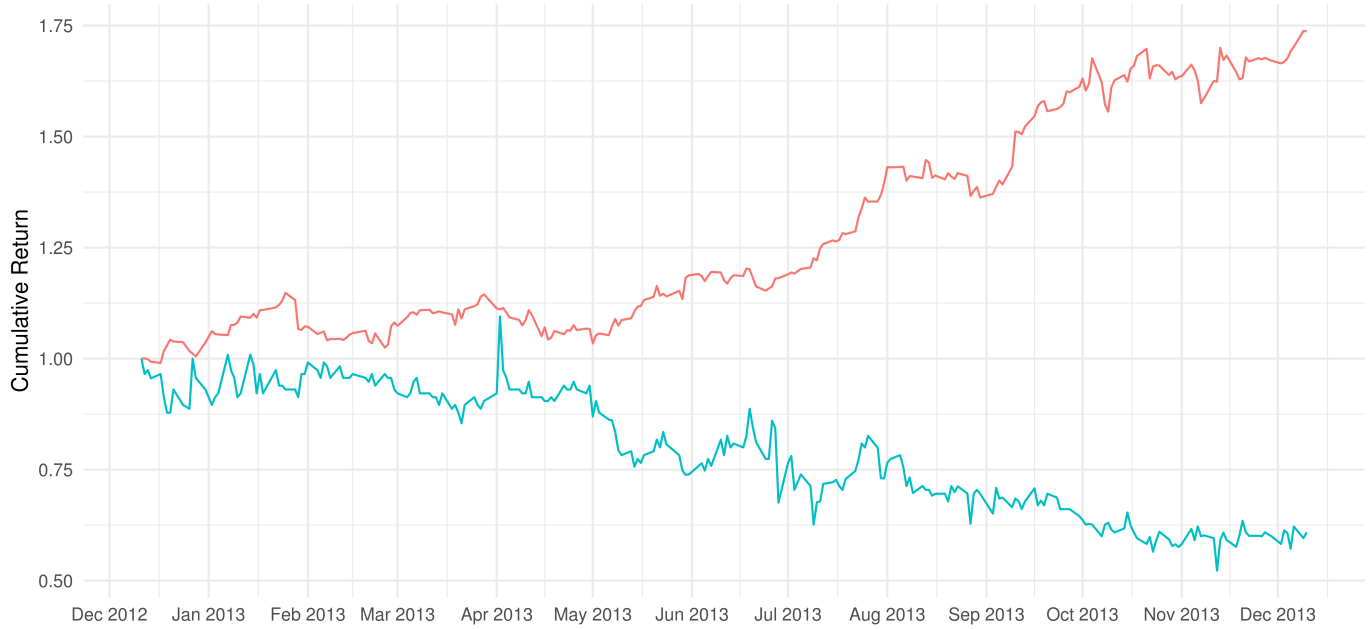


Step 2: Generate all pairs of stocks on each day

Step 3: Predict if stock ABC will outperform stock XYZ over the next year

Classification Model

- Training set contains 1 day every 3 months (expanding)
- If we don't use enough data, our model may "memorize" the best/worst stocks

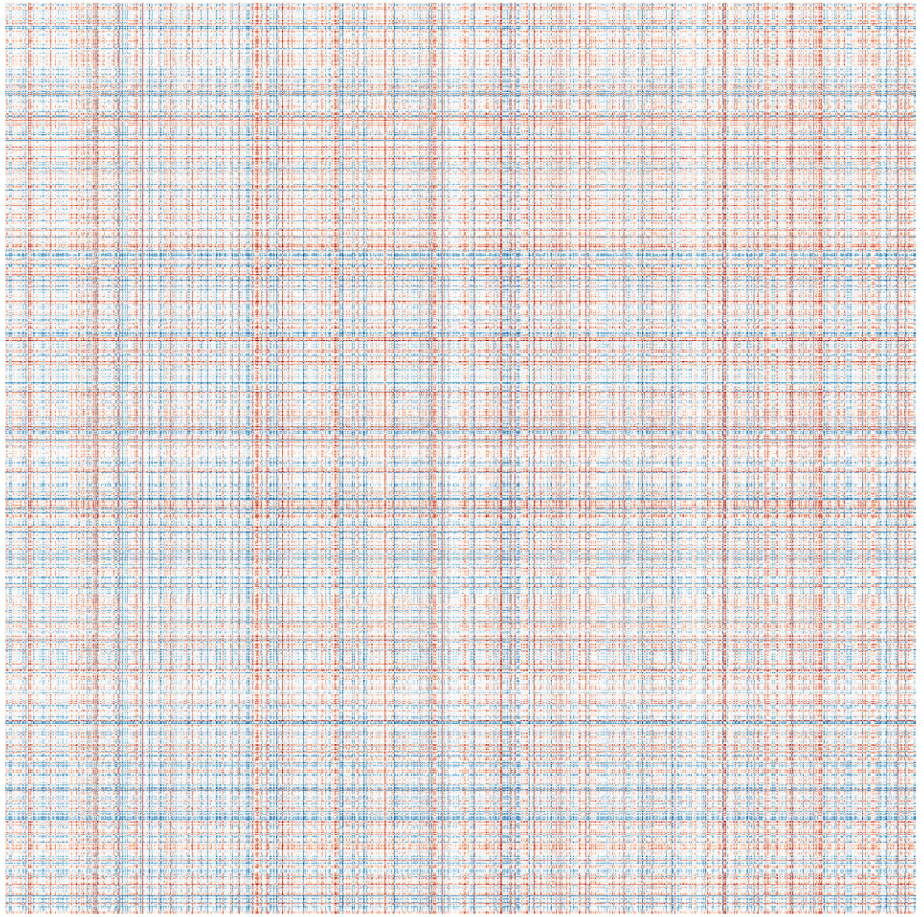


Step 4: Generate a probability matrix

A matrix that contains the predicted probability of stock j outperforming stock i over the next year:

$$\mathbf{P}_{i,j} = \Pr(X_i \leq X_j)$$

$$\mathbf{P} =$$



Step 5: Probability of each stock outperforming/underperforming all other stocks

Probability of stock i outperforming all other stocks:

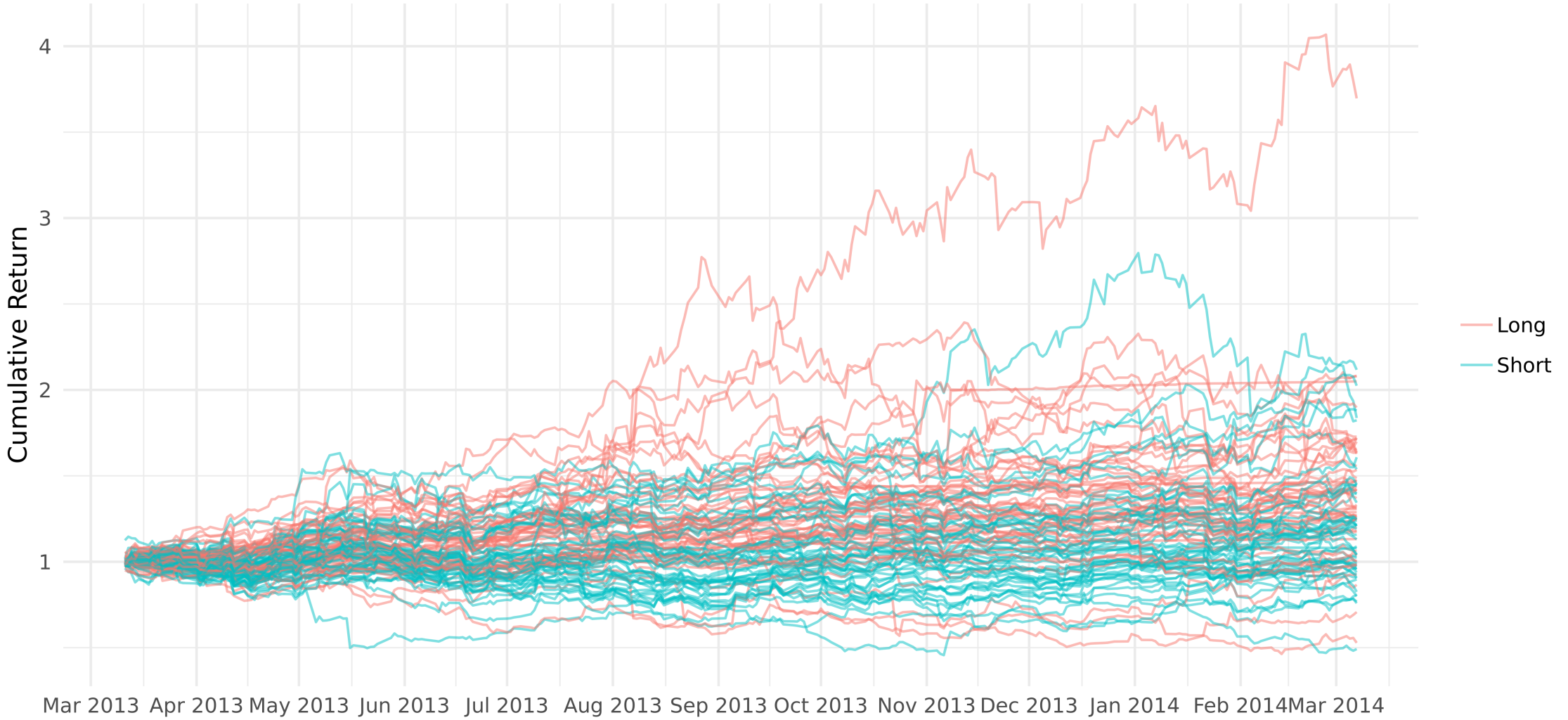
$$\Pr(X_i > X_1 \,\&\, X_i > X_2 \,\&\, \cdots \,\&\, X_i > X_N) = \prod_{\substack{j=1 \\ j \neq i}}^{N} 1 - \Pr(X_i \leq X_j)$$

Probability of stock i underperforming all other stocks:

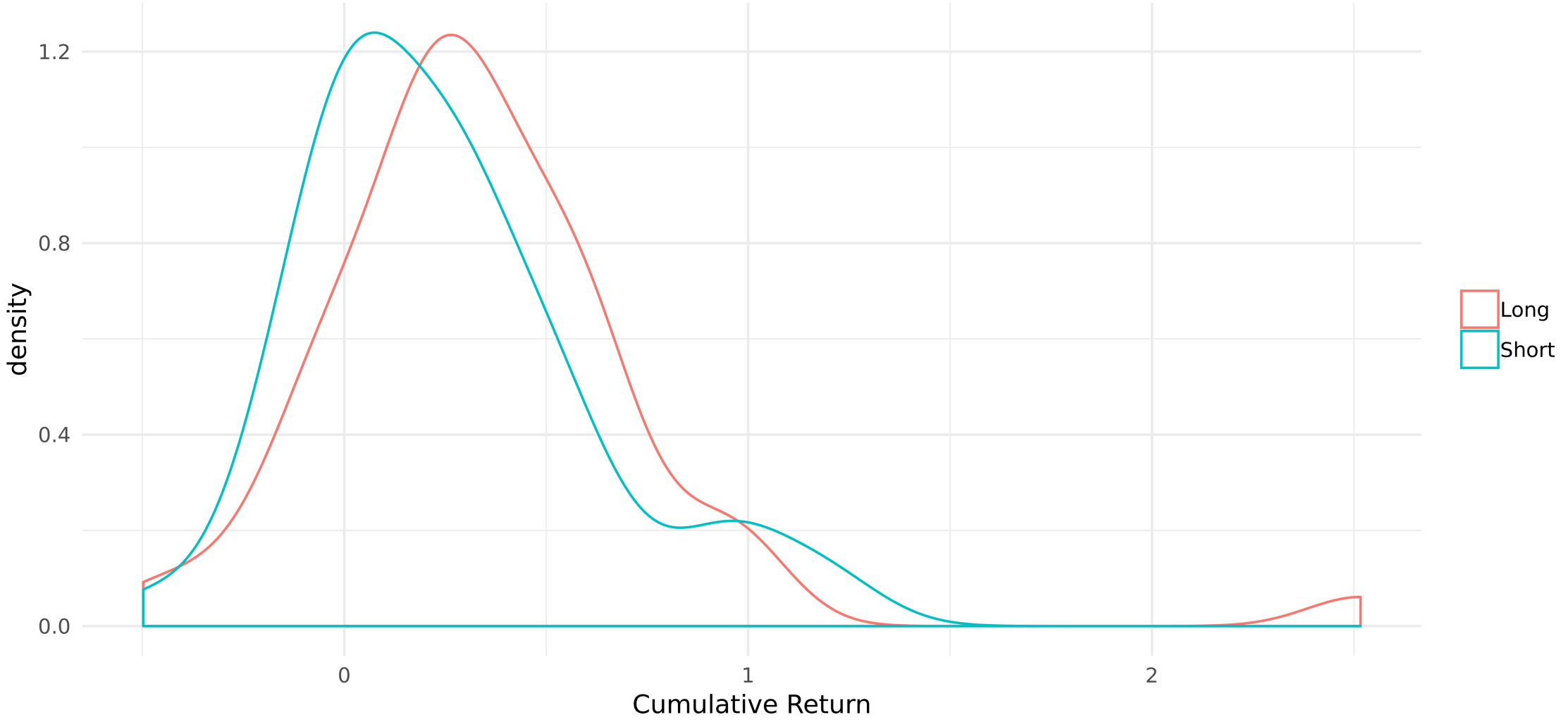
$$\Pr(X_i \leq X_1 \,\&\, X_i \leq X_2 \,\&\, \cdots \,\&\, X_i \leq X_N) = \prod_{\substack{j=1 \\ j \neq i}}^{N} \Pr(X_i \leq X_j)$$

Stock	Log Probability of Best	Log Probability of Worst
Stock 1	-478.86101	-1311.6859
Stock 2	-1033.0009	-605.94667
Stock 3	-740.48224	-841.6922
Stock 4	-868.96107	-713.01704
Stock 5	-666.08427	-911.40809

Step 6: Long the top 50 stocks each day, short the bottom 50



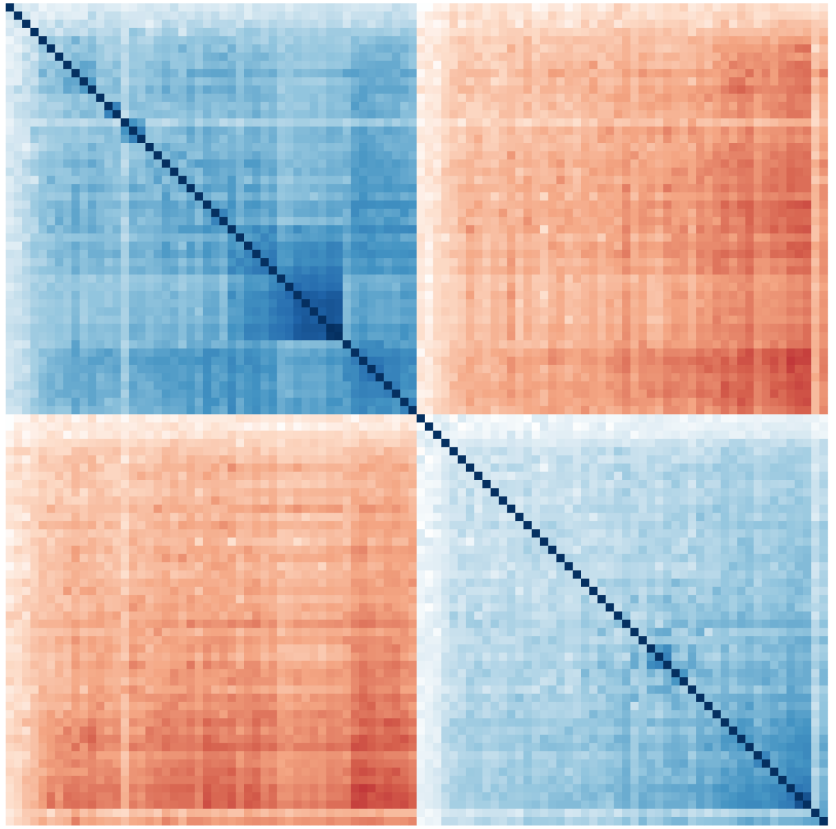
Step 6: Long the top 50 stocks each day, short the bottom 50



Step 7: Portfolio construction algorithm each day to determine stock weights

Overview of Algorithm

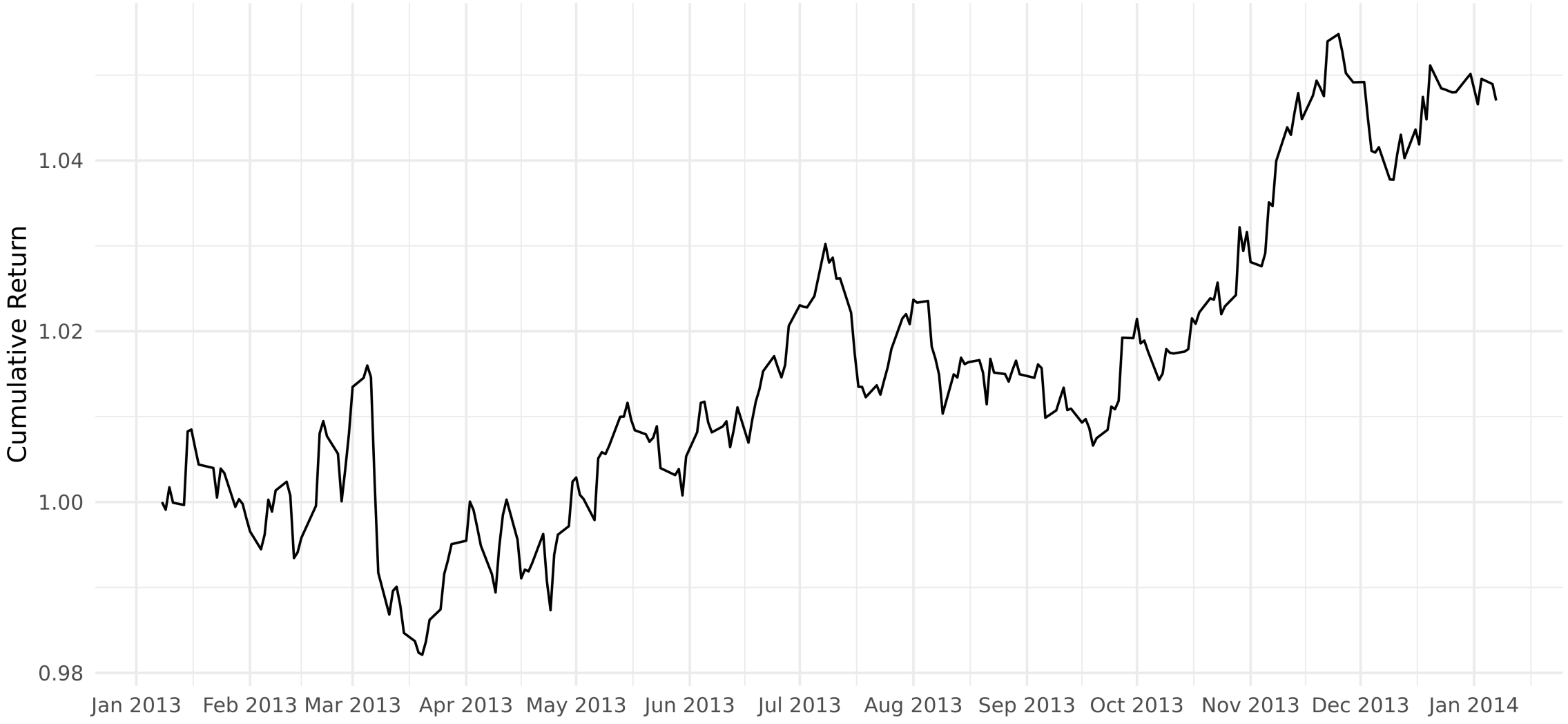
1. Cluster the portfolios using hierarchical clustering (single linkage)
2. Sort based on the clustering
3. Split portfolios in half, weighting both halves by their inverse portfolio volatility



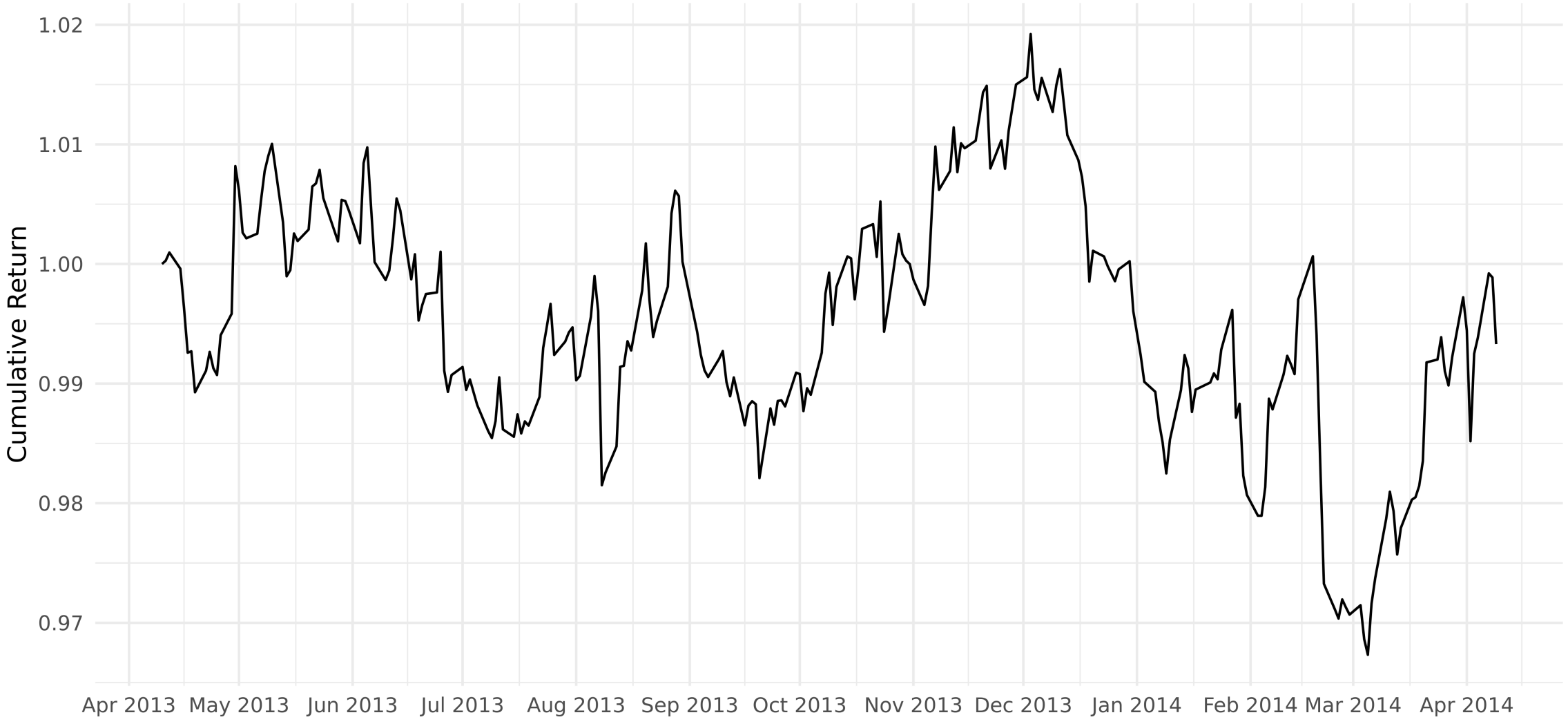
López de Prado, Marcos, Building Diversified Portfolios that Outperform Out-of-Sample (May 23, 2016). Journal of Portfolio Management, 2016, Forthcoming.

Step 7: Portfolio construction algorithm each day to determine stock weights

Step 7: Portfolio construction algorithm each day to determine stock weights



Step 7: Portfolio construction algorithm each day to determine stock weights



Step 7: Portfolio construction algorithm each day to determine stock weights

Step 8: Layer portfolios each day

Questions?