Stock Price Prediction and Portfolio Optimization Using Recurrent Neural Networks and Autoencoders

Predictive Analytics World 2020 Berlin

Julian Quernheim

Introduction

- Methods for calculating an optimal stock portfolio still **focus on historic data of stock returns** and additionally have **difficulties capturing non-linearities** of the timeseries when quantifying risk.

- This talk addresses this problem using deep learning methods.

- A basic knowledge of deep learning and portfolio management is required.

Introduction

Focus, forecast and clean.

1 Which stocks to analyze?

- **Focus on stocks that move the market to decrease computation time!**

2 Does forecasting improve the portfolio?

- **Don’t forecast too far. A forecast is only a strong indicator.**

3 How to improve the risk calculation of a stock?

- **Try to capture non-linearities in the time series.**

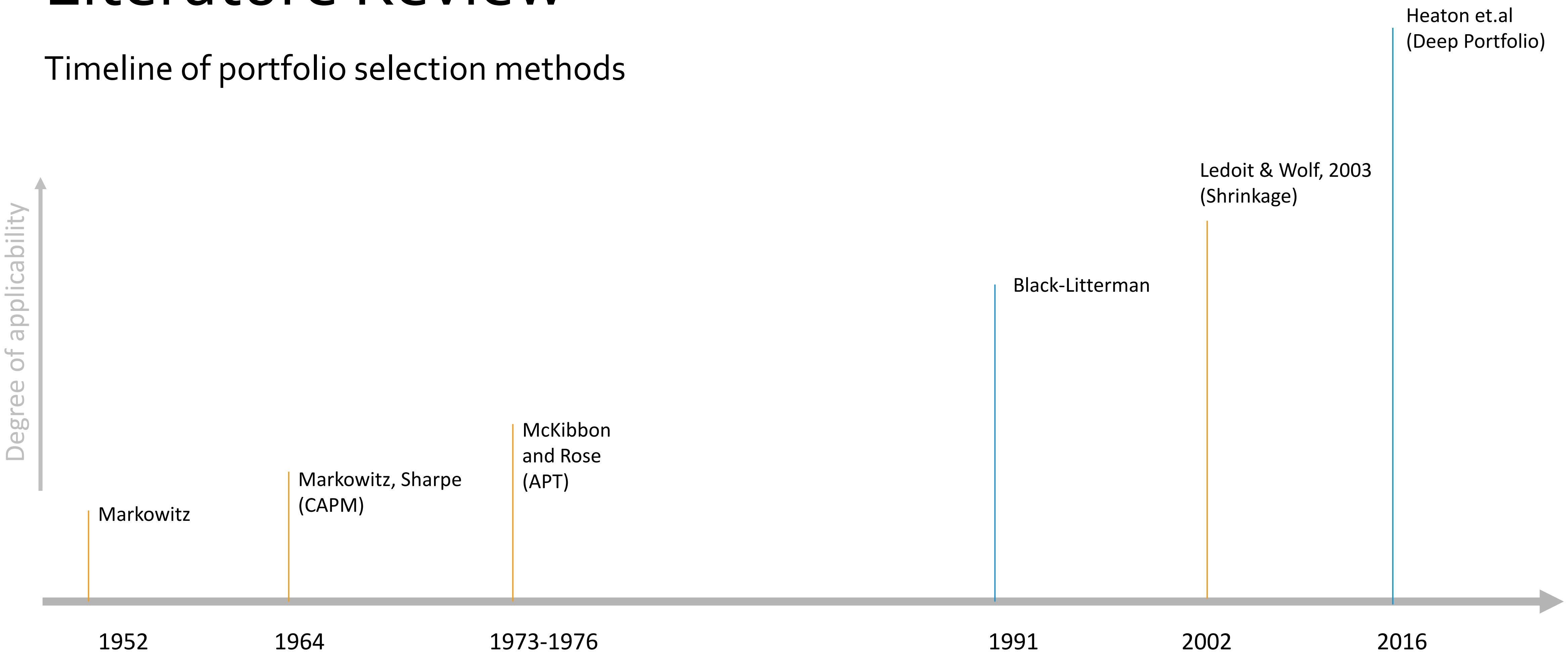
4 How to calculate an optimal portfolio?

- **Don’t trust the in-sample results. Look at the out-of-sample results.**

Literature Review

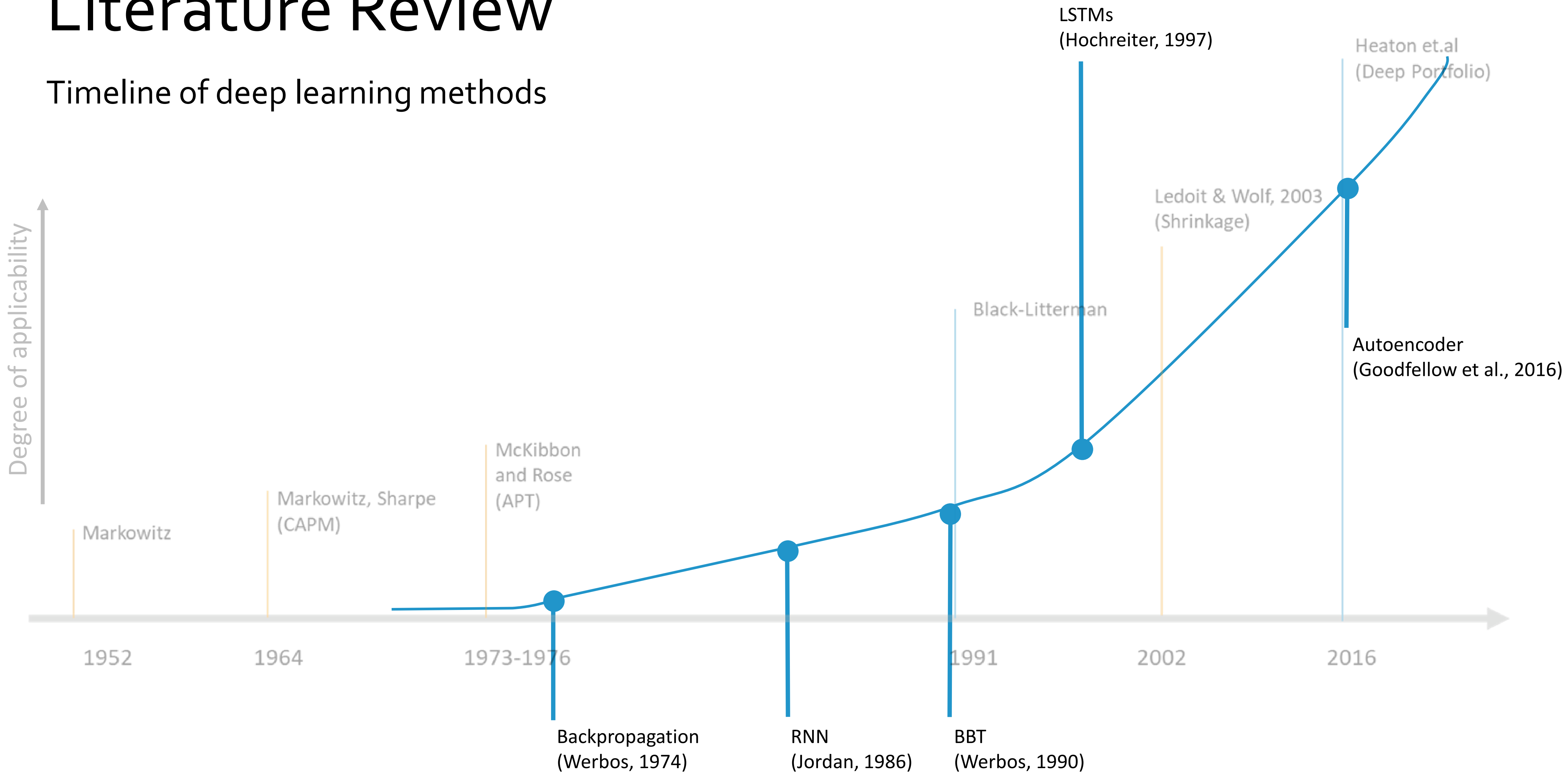
Literature Review

Timeline of portfolio selection methods



Literature Review

Timeline of deep learning methods



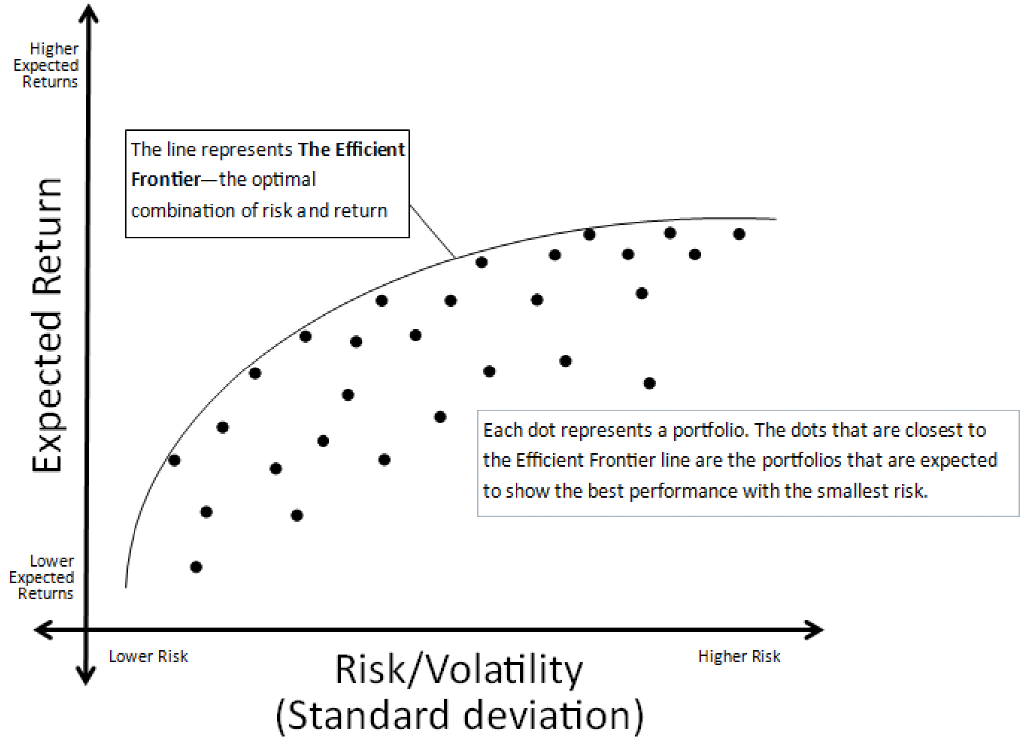
Timeline of deep learning methods

Literature Review

Modern portfolio theory (MPT) (Markowitz, 1952)

What is MPT?

- Investor can construct a portfolio of multiple assets that will **maximize returns** (r_i) for a given level of portfolio risk.
- Likewise, given a desired level of expected return, an investor can construct a portfolio that **minimizes risk**.



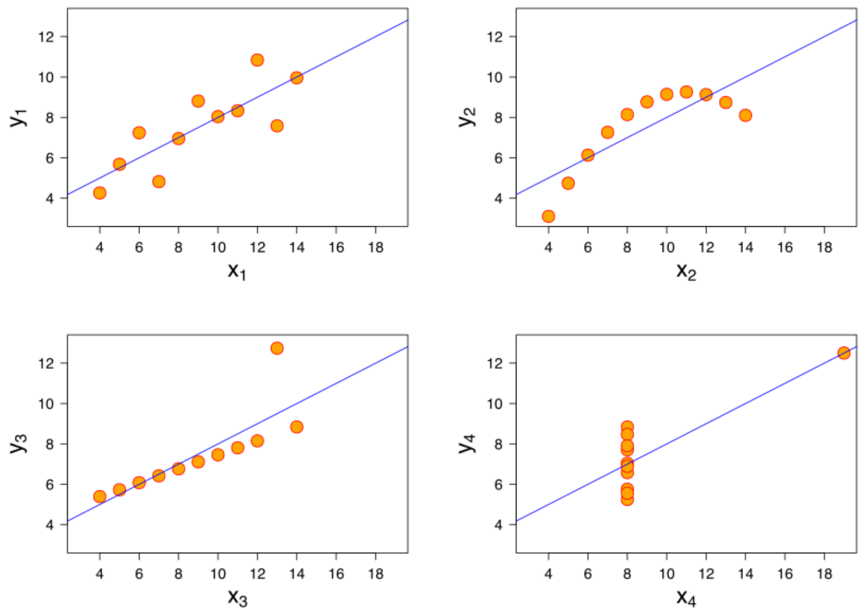
Efficient frontier: Source: http://www.alamedafinancialgroup.com/Our-Investment-Phiosophy.5.htm

Literature Review

Modern portfolio theory (MPT) (Markowitz, 1952)

Pitfalls of covariance

- The covariance indicates a **linear** relationship between two variables. Hence it can be fallacious in situations where two variables have a relationship, but it is **nonlinear**.
- Covariance is strongly influenced by **outliers**.



Anscombe's quartet: All four sets are identical when examined using simple summary statistics but vary considerably when graphed.

Solutions

- Shrinkage Methods (Ledoit & Wolf, 2003)

$$\hat{C} = \gamma C + (1 - \gamma) * B$$

where:

- B is the shrinkage estimator and
- γ is the shrinkage constant

mVISE

Literature Review

Modern portfolio theory (MPT) (Markowitz, 1952)

Linear programming formulation

$$\text{minimize } C\, w^T w$$

$$s.t. \qquad w^T \mu \geq \mu_b$$

$$w^T \mathbf{1} = 1$$

$$w_i \geq 0$$

Where:

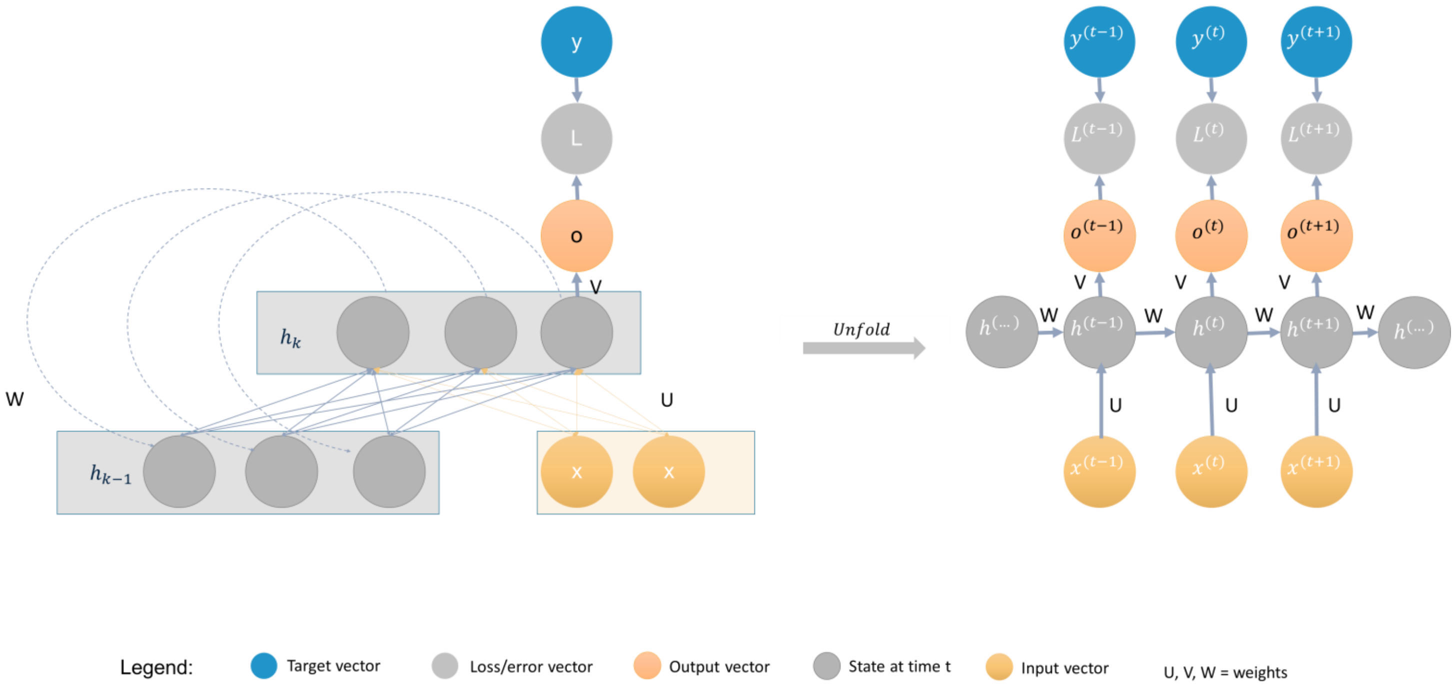
- the return of stock i ∈ {1–n} is defined as r_i
- the expected return vector is defined as $\mu = \begin{pmatrix} E(r_1) \\ \vdots \\ E(r_n) \end{pmatrix}$, μ_b is the market return.
- the weights of the stocks in a portfolio equals $\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$
- the covariance of two stock returns 1 and 2 equals $cov(r_1, r_2)$, *and* the covariance matrix of all stocks is n is $cov(\boldsymbol{r}, \boldsymbol{r}) = C$

Source: https://medium.com/datadriveninvestor/simple-portfolio-optimization-harry-markowitz-mean-variance-model-using-excel-part-1-efc3f19a347

Literature Review

How do Recurrent neural networks work?

- Recurrent neural networks (RNNs) **process sequential data**.
- Each state of the RNN is therefore a function depending on its previous states.



An example of a folded and unfolded RNN.

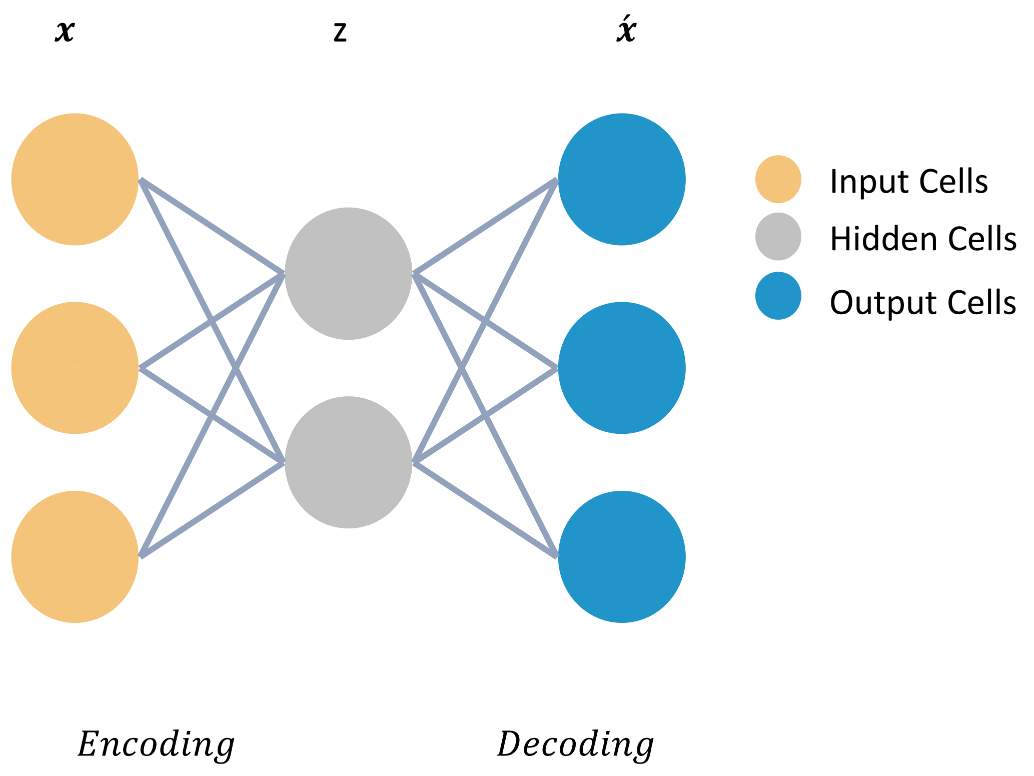
Literature Review

What are Autoencoders?

An autoencoder is an unsupervised neural network that is trained to attempt to copy its input to its output [Goodfellow et al., 2016].

Autoencoders are used for:

- dimensionality reduction
- removing structural noise
- feature learning
- outlier detection



Example of an undercomplete autoencoder with three input and output layers and two hidden layers.

Data, Methodology & Results

Data, Methodology & Results

1 **Which stocks to analyze?**
→ Apply an autoencoder model and filter stocks that can be recreated best

2 **Does forecasting improve the portfolio?**
→ Forecast the next 10-days of a stock closing value into the future using Recurrent Neural networks

3 **How to improve the risk calculation of a stock?**
→ Apply latent features of an autoencoder model to clean the sample covariance matrix

4 **How to calculate an optimal portfolio?**
→ Apply Markowitz portfolio optimization and find the optimal stocks for the portfolio

Dataset

- Dataset: daily-historical-stock-prices-1970-2018
- Source: Kaggle
- Stock exchanges NYSE and NASDAQ
- Dataset dimension: [20973889 rows x 8 columns]
- Tickers: 5685

	ticker	open	close	adj_close	low	high	volume	date
0	AHH	11.50000	11.58000	8.49315	11.25000	11.68000	4633900	2013-05-08
1	AHH	11.66000	11.55000	8.47115	11.50000	11.66000	275800	2013-05-09
2	AHH	11.55000	11.60000	8.50782	11.50000	11.60000	277100	2013-05-10
3	AHH	11.63000	11.65000	8.54449	11.55000	11.65000	147400	2013-05-13
4	AHH	11.60000	11.53000	8.45648	11.50000	11.60000	184100	2013-05-14
5	AHH	11.60000	11.60000	8.50782	11.54000	11.60000	76800	2013-05-15
6	AHH	11.62000	11.74000	8.61050	11.54000	11.74000	170300	2013-05-16
7	AHH	11.70000	11.76000	8.62517	11.70000	11.85000	305400	2013-05-17
8	AHH	11.76000	11.73000	8.60317	11.63000	11.83000	46800	2013-05-20
9	AHH	11.76000	11.83000	8.67651	11.61000	11.84000	77000	2013-05-21
10	AHH	11.84000	11.75000	8.61784	11.68000	11.84000	90200	2013-05-22
11	AHH	11.64000	11.61000	8.51516	11.54000	11.79000	75400	2013-05-23
12	AHH	11.57000	11.70000	8.58117	11.55000	11.73000	63400	2013-05-24
13	AHH	11.73000	11.58000	8.49315	11.57000	11.78000	212400	2013-05-28
14	AHH	11.51000	11.25000	8.25112	11.20000	11.67000	447000	2013-05-29
15	AHH	11.58000	11.55000	8.47115	11.47000	11.65000	453300	2013-05-30
16	AHH	11.60000	11.47000	8.41248	11.47000	11.80000	109400	2013-05-31
17	AHH	11.68000	11.52000	8.44915	11.30000	11.68000	143200	2013-06-03
18	AHH	11.56000	11.51000	8.44182	11.46000	11.63000	115000	2013-06-04
19	AHH	11.58000	11.55000	8.47115	11.25000	11.59000	78100	2013-06-05
20	AHH	11.51000	11.62000	8.52249	11.50000	11.70000	121700	2013-06-06
21	AHH	11.67000	11.60000	8.50782	11.52000	11.68000	52300	2013-06-07
22	AHH	11.60000	11.65000	8.54449	11.50000	11.67000	167200	2013-06-10
23	AHH	11.56000	11.56000	8.47849	11.35000	11.60000	117500	2013-06-11
24	AHH	11.55000	11.49000	8.42715	11.45000	11.56000	162100	2013-06-12
25	AHH	11.48000	11.49000	8.42715	11.30000	11.55000	63600	2013-06-13
26	AHH	11.49000	11.45000	8.39781	11.30000	11.49000	73100	2013-06-14
27	AHH	11.48000	11.55000	8.47115	11.40000	11.65000	73300	2013-06-17
28	AHH	11.55000	11.51000	8.44182	11.39000	11.60000	149700	2013-06-18
29	AHH	11.48000	11.30000	8.28780	11.22000	11.48000	80300	2013-06-19
30	AHH	11.25000	11.13000	8.16311	11.13000	11.30000	203100	2013-06-20
31	AHH	11.17000	11.27000	8.26579	11.12000	11.27000	287900	2013-06-21
32	AHH	11.20000	11.22000	8.22912	10.70000	11.29000	151500	2013-06-24
33	AHH	11.28000	11.38000	8.34647	10.98000	11.40000	95600	2013-06-25
34	AHH	11.44000	11.38000	8.34647	10.90000	11.46000	173900	2013-06-26
35	AHH	11.33000	11.46000	8.46465	11.20000	11.54000	153300	2013-06-27
36	AHH	11.43000	11.78000	8.70101	11.37000	11.78000	1659800	2013-06-28
37	AHH	11.74000	11.67000	8.61976	11.41000	11.84000	103800	2013-07-01
38	AHH	11.70000	11.71000	8.64930	11.45000	11.80000	98600	2013-07-02
39	AHH	11.69000	11.74000	8.67146	11.58000	11.75000	45100	2013-07-03
40	AHH	11.84000	11.82000	8.73055	11.50000	11.84000	89700	2013-07-05
41	AHH	11.75000	11.47000	8.47203	11.35000	11.75000	170100	2013-07-08
42	AHH	11.52000	11.19000	8.26522	11.10000	11.53000	242600	2013-07-09
43	AHH	11.15000	11.03000	8.14704	10.55000	11.15000	206800	2013-07-10
44	AHH	11.12000	11.14000	8.22829	10.99000	11.16000	189100	2013-07-11

Original stock dataset

Dataset

- Selected last 1000 days from 2014 -2018
- Transformed each ticker into columns
- Final dataset: [1000 rows x 13925 columns]

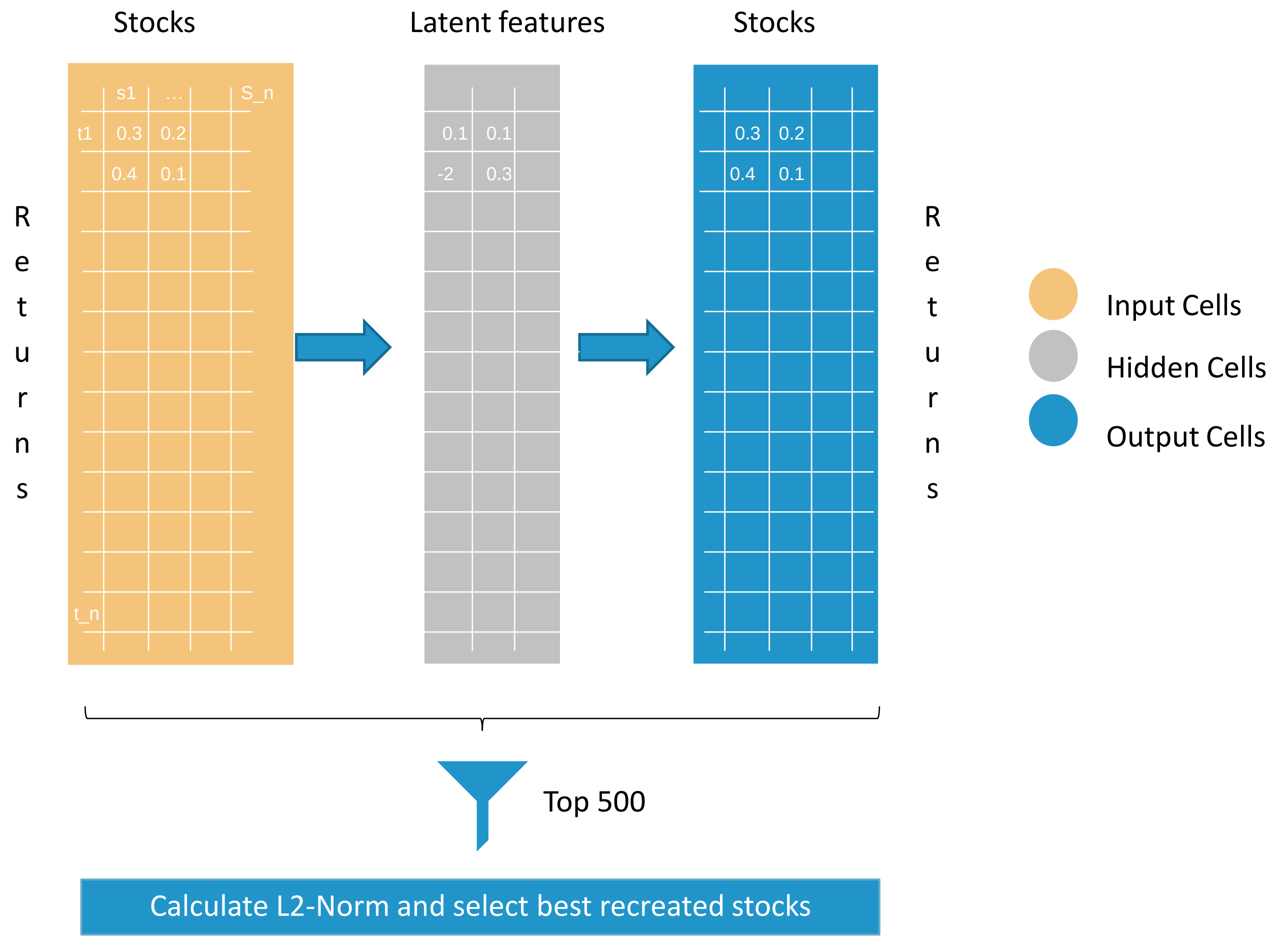
	KOOL_Open	KOOL_High	KOOL_Low	KOOL_Close	KOOL_Volume	ADXS_Open	ADXS_High	ADXS_Low	ADXS_Close	ADXS_Volume	ACHV_Open	ACHV_High	ACHV_Low	ACHV_Close	ACHV_Volume
2018-07-13T00:00:00.000000000	0.47000	0.51000	0.44000	0.46000	414300.00000	1.60000	1.62000	1.30000	1.34000	6967800.00000	3.62000	3.74000	3.41000	3.55000	354400.00000
2018-07-16T00:00:00.000000000	0.51000	0.52000	0.47000	0.48000	1027000.00000	1.34000	1.51000	1.31000	1.44000	2091200.00000	3.53000	3.61000	3.41000	3.45000	264300.00000
2018-07-17T00:00:00.000000000	0.50000	0.50000	0.43000	0.43000	843200.00000	1.44000	1.47500	1.36000	1.41000	1318600.00000	3.42000	3.50000	3.35000	3.36000	175300.00000
2018-07-18T00:00:00.000000000	0.43000	0.45000	0.42000	0.43000	301500.00000	1.40000	1.46000	1.37000	1.43000	560700.00000	3.36000	3.41000	3.25000	3.33000	170700.00000
2018-07-19T00:00:00.000000000	0.45000	0.45000	0.41000	0.42000	219400.00000	1.44000	1.49000	1.41100	1.45000	693500.00000	3.30000	3.38000	3.25000	3.35000	119600.00000
2018-07-20T00:00:00.000000000	0.42000	0.43000	0.42000	0.42000	85800.00000	1.46000	1.49000	1.42000	1.48000	533700.00000	3.34000	3.43000	3.26000	3.30000	79900.00000
2018-07-23T00:00:00.000000000	0.42000	0.43000	0.41000	0.42000	87700.00000	1.48000	1.62000	1.47000	1.59000	978100.00000	3.28000	3.50000	3.21000	3.39000	136800.00000
2018-07-24T00:00:00.000000000	0.42000	0.43000	0.41000	0.42000	65300.00000	1.62000	1.64000	1.48000	1.51000	568000.00000	3.41000	3.46000	3.31000	3.36000	62700.00000
2018-07-25T00:00:00.000000000	0.42000	0.42000	0.41000	0.42000	158500.00000	1.51000	1.52000	1.46100	1.50000	201700.00000	3.37000	3.39000	3.26000	3.31000	60400.00000
2018-07-26T00:00:00.000000000	0.42000	0.42000	0.39000	0.40000	285400.00000	1.48000	1.50000	1.38000	1.40000	663700.00000	3.31000	3.42000	3.26000	3.37000	57800.00000
2018-07-27T00:00:00.000000000	0.42000	0.42000	0.39000	0.40000	169800.00000	1.40000	1.43000	1.36000	1.38000	452200.00000	3.39000	3.44000	3.25000	3.28000	83300.00000
2018-07-30T00:00:00.000000000	0.41000	0.42000	0.38000	0.40000	108200.00000	1.50000	1.50000	1.40000	1.43000	369900.00000	3.26000	3.43000	3.25000	3.31000	51800.00000
2018-07-31T00:00:00.000000000	0.40000	0.41000	0.38000	0.38000	120800.00000	1.42000	1.48000	1.41000	1.46000	332400.00000	3.38000	3.38000	3.25000	3.28000	87100.00000
2018-08-01T00:00:00.000000000	0.37000	0.39000	0.36000	0.38000	148100.00000	1.45000	1.49000	1.41000	1.43000	134400.00000	3.30000	3.30000	3.07000	3.22000	95400.00000
2018-08-02T00:00:00.000000000	0.38000	0.40000	0.37000	0.38000	101200.00000	1.42000	1.44000	1.33000	1.38000	422300.00000	3.17000	3.22000	3.11000	3.22000	54600.00000
2018-08-03T00:00:00.000000000	0.37000	0.40000	0.37000	0.39000	67000.00000	1.42000	1.42000	1.35000	1.41000	205000.00000	3.23000	3.30000	3.15000	3.18000	42000.00000
2018-08-06T00:00:00.000000000	0.39000	0.40000	0.37000	0.38000	92800.00000	1.41000	1.45000	1.37400	1.43000	216300.00000	3.22000	3.30000	3.20000	3.29000	30800.00000
2018-08-07T00:00:00.000000000	0.40000	0.40000	0.37000	0.38000	186500.00000	1.42000	1.44000	1.40000	1.44000	145900.00000	3.32000	3.48000	3.32000	3.46000	147300.00000
2018-08-08T00:00:00.000000000	0.39000	0.39000	0.29000	0.32000	741200.00000	1.44000	1.51000	1.40000	1.51000	392100.00000	3.45000	3.48000	3.30000	3.42000	49900.00000
2018-08-09T00:00:00.000000000	0.33000	0.34000	0.30000	0.32000	159700.00000	1.50000	1.54000	1.45000	1.45000	315500.00000	3.36000	3.40000	3.10000	3.12000	132600.00000
2018-08-10T00:00:00.000000000	0.34000	0.35000	0.31000	0.34000	292100.00000	1.45000	1.49000	1.42000	1.45000	134100.00000	3.10000	3.12000	2.87000	3.08000	191900.00000
2018-08-13T00:00:00.000000000	0.36000	0.36000	0.30000	0.34000	230900.00000	1.45000	1.48000	1.41000	1.48000	238900.00000	3.07000	3.17000	2.92000	3.05000	73500.00000
2018-08-14T00:00:00.000000000	0.38000	0.38000	0.28000	0.28000	416700.00000	1.47000	1.52000	1.46100	1.51000	269100.00000	3.01000	3.07000	2.90000	3.05000	56100.00000
2018-08-15T00:00:00.000000000	0.32000	0.32000	0.26000	0.29000	647300.00000	1.51000	1.51000	1.42000	1.43000	212600.00000	2.93000	3.07000	2.67000	2.85000	78500.00000
2018-08-16T00:00:00.000000000	0.30000	0.30000	0.27000	0.28000	165500.00000	1.43000	1.47000	1.40000	1.44000	141000.00000	2.91000	2.98000	2.71000	2.84000	53000.00000
2018-08-17T00:00:00.000000000	0.28000	0.28000	0.27000	0.28000	81700.00000	1.44000	1.45000	1.41000	1.44000	109200.00000	2.86000	2.89000	2.71000	2.80000	44900.00000
2018-08-20T00:00:00.000000000	0.28000	0.28000	0.22000	0.25000	830900.00000	1.43000	1.45000	1.40000	1.43000	143000.00000	2.94000	3.26000	2.91000	3.04000	126600.00000
2018-08-21T00:00:00.000000000	0.26000	0.27000	0.23000	0.27000	416400.00000	1.42000	1.47000	1.40000	1.42000	245500.00000	3.08000	3.29000	3.06000	3.24000	140800.00000

Transformed stock dataset

mVISE

1 Which stocks to analyze?

Focus on stocks that move the market!



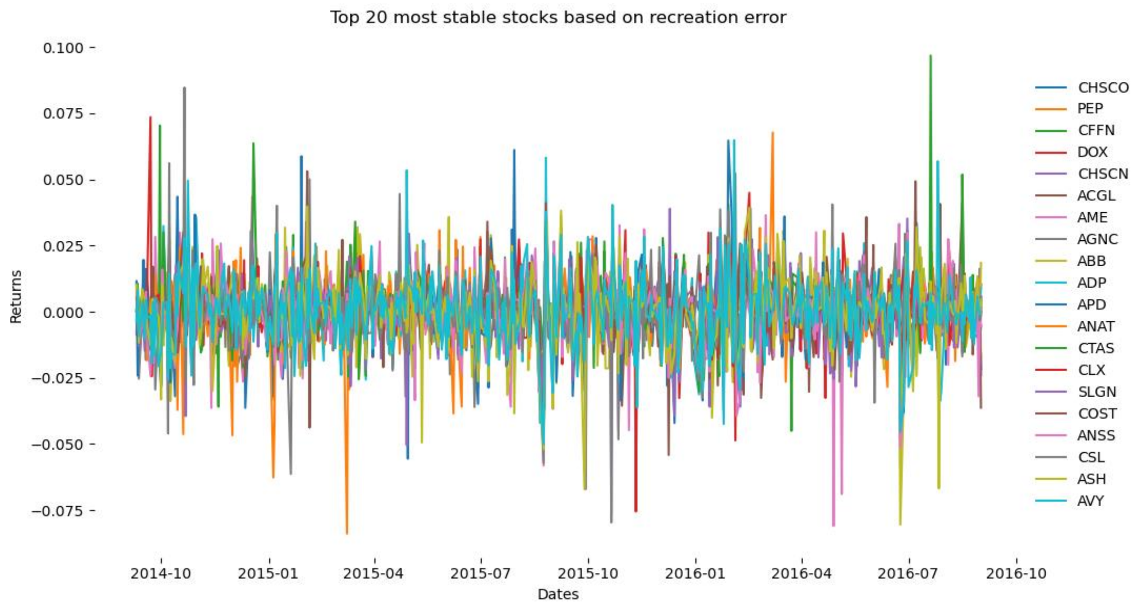
Autoencoder model with ranked recreation error.

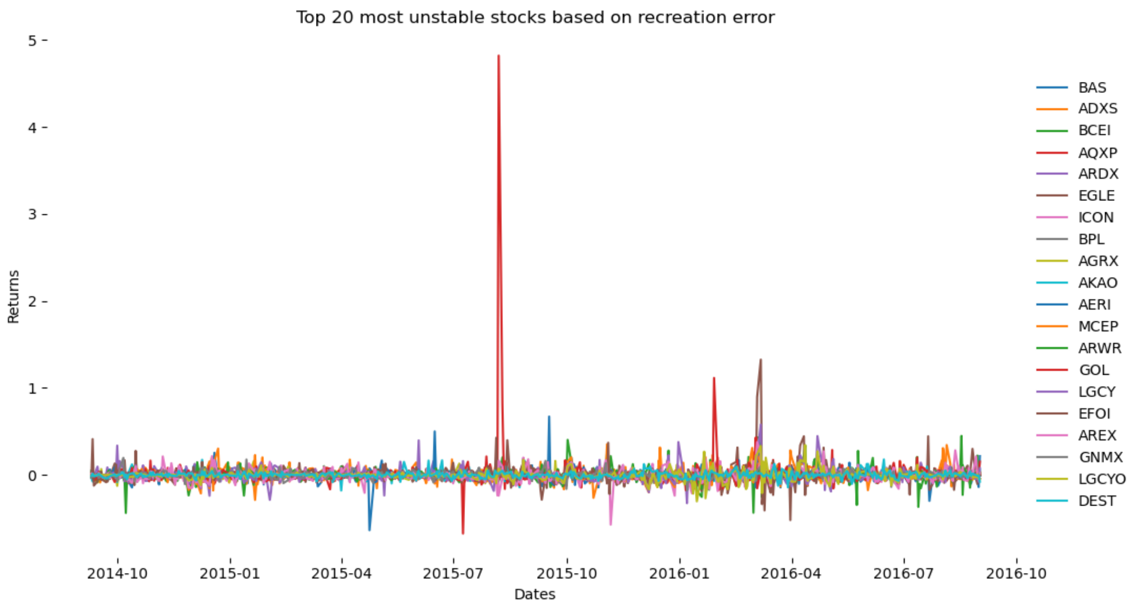
Intuition:

The stocks with the lowest recreation error (L2-norm) represent the market better. They are less volatile and are considered to be similar to large cap stocks.

Which stocks to analyze?

Unstable stocks tend to be more volatile and have more unexpected spikes!

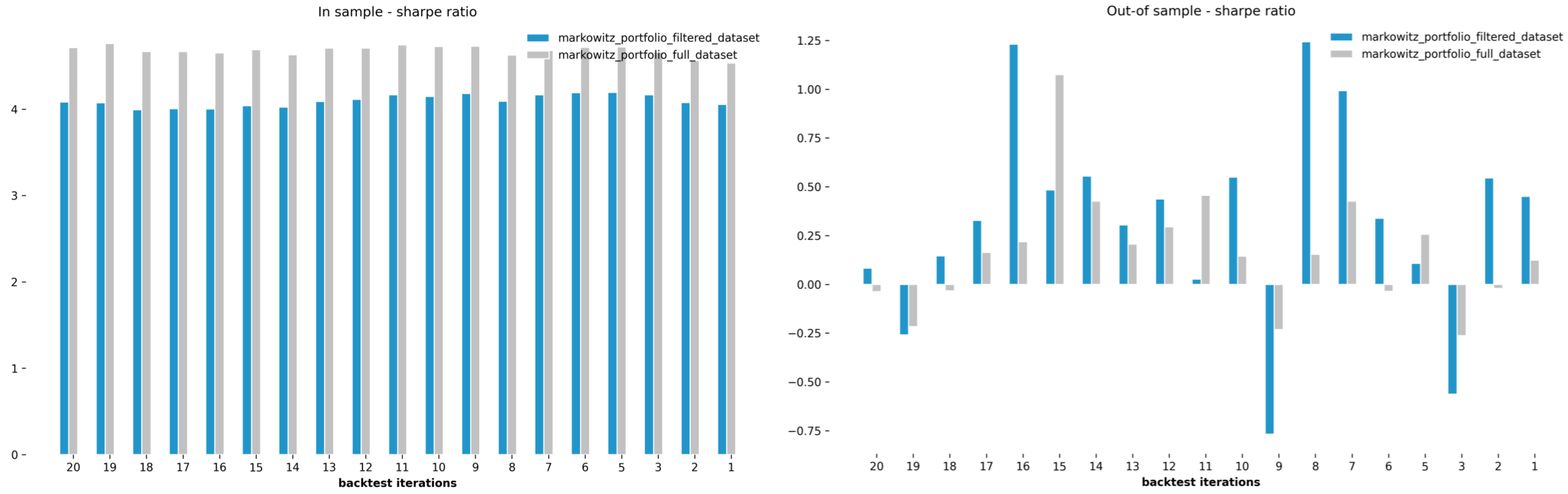




Top 20 most stable and most unstable stocks ranked by recreation error.

1 Which stocks to analyze?

Filtering based on recreation error improves out-of sample performance!



In-sample and out-of-sample sharpe ratio of models trained on full dataset and filtered dataset.

mVISE

2 Does forecasting improve the portfolio?

Data transformation is still part of the job!

	GOOGL_Open	GOOGL_High	GOOGL_Low	GOOGL_Close	GOOGL_Volume
2014-09-08	599.14001	603.52002	598.02002	601.63000	1599200.00000
2014-09-09	600.27002	600.59998	590.47998	591.96997	1571200.00000
2014-09-10	591.73999	593.71997	587.14001	593.41998	1160100.00000
2014-09-11	590.00000	591.95001	586.03998	591.10999	1501400.00000
2014-09-12	590.39001	591.47998	583.28003	584.90002	1851500.00000
2014-09-15	582.23999	583.91998	577.01001	581.64001	1545400.00000
2014-09-16	580.95001	590.15002	580.95001	588.78003	1579600.00000
2014-09-17	589.51001	596.07001	587.12000	593.28998	1719500.00000
2014-09-18	595.04999	597.56000	593.02002	597.27002	1494500.00000
2014-09-19	599.48999	605.40002	597.76001	605.40002	4191600.00000
2014-09-22	602.50000	603.79999	593.12000	597.27002	1782200.00000
2014-09-23	595.00000	596.65002	590.23999	591.17999	1403700.00000
2014-09-24	591.57001	600.10999	590.29999	598.41998	1758000.00000
2014-09-25	596.98999	598.07001	584.73999	585.25000	1671000.00000
2014-09-26	585.92999	589.57001	585.00000	587.90002	1289100.00000
2014-09-29	581.83002	589.31000	581.57001	587.81000	1143700.00000
2014-09-30	587.48999	591.00000	584.50000	588.40997	1571500.00000
2014-10-01	586.79999	588.71997	578.02002	579.63000	1447700.00000
2014-10-02	578.00000	583.23999	574.04999	580.88000	1536300.00000
2014-10-03	584.09998	588.28998	583.50000	586.25000	1214500.00000
2014-10-06	589.95001	592.40002	585.40002	587.78003	1286100.00000
2014-10-07	584.90002	585.84998	574.09998	574.09998	1545700.00000
2014-10-08	574.78998	584.69000	567.64001	583.73999	2207900.00000
2014-10-09	581.60999	582.53003	569.03003	570.81000	2411700.00000
2014-10-10	567.46997	575.22998	555.01001	555.19000	2978700.00000
2014-10-13	555.13000	560.88000	544.42999	544.75000	2755800.00000
2014-10-14	550.14001	558.63000	544.50000	548.69000	2609900.00000
2014-10-15	542.08002	543.91998	528.41998	540.72998	3836000.00000
2014-10-16	527.00000	540.98999	524.95001	536.91998	3805000.00000
2014-10-17	540.45001	543.45001	518.40997	522.96997	5996500.00000
2014-10-20	520.45001	533.15997	519.14001	532.38000	2748200.00000
2014-10-21	537.27002	538.77002	530.20001	538.03003	2459500.00000

50 days sliding window

```
array([[[0.12647494, 0.13049197, 0.13872737, 0.132629  , 0.08742452],
        [0.12790554, 0.12680113, 0.12896162, 0.12037691, 0.08515502],
        [0.1171064 , 0.11810503, 0.12463574, 0.122216  , 0.05183384],
        ...,
        [0.07837904, 0.07706406, 0.08365605, 0.07760895, 0.0464843 ],
        [0.07794856, 0.07652057, 0.08182987, 0.07531328, 0.08027558],
        [0.07439106, 0.07295617, 0.07978345, 0.07372787, 0.07331307]],

       [[0.12790554, 0.12680113, 0.12896162, 0.12037691, 0.08515502],
        [0.1171064 , 0.11810503, 0.12463574, 0.122216  , 0.05183384],
        [0.11490354, 0.11586787, 0.12321099, 0.11928617, 0.07949747],
        ...,
        [0.07794856, 0.07652057, 0.08182987, 0.07531328, 0.08027558],
        [0.07439106, 0.07295617, 0.07978345, 0.07372787, 0.07331307],
        [0.07083356, 0.06940443, 0.06933129, 0.06288369, 0.09402229]],

       [[0.1171064 , 0.11810503, 0.12463574, 0.122216  , 0.05183384],
```

50 day sliding window example for RNN model.

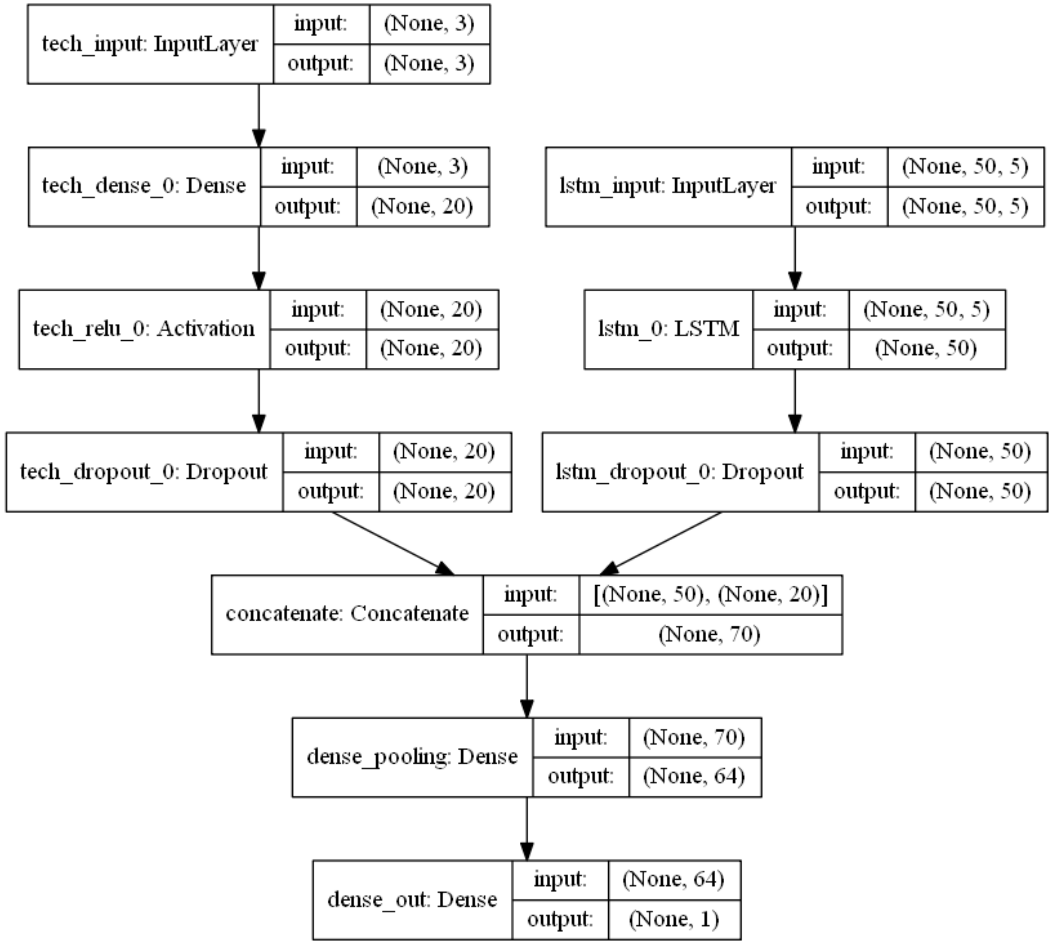
2 Does forecasting improve the portfolio?

Using a multi-input model is a good way to improve accuracy!

Model Design:

A multi-input model has been applied using Keras functional API, to include:

- historic stock prices (ohlcv)
- additional technical indicators e.g. exponential moving average



Keras RNN model.

mVISE

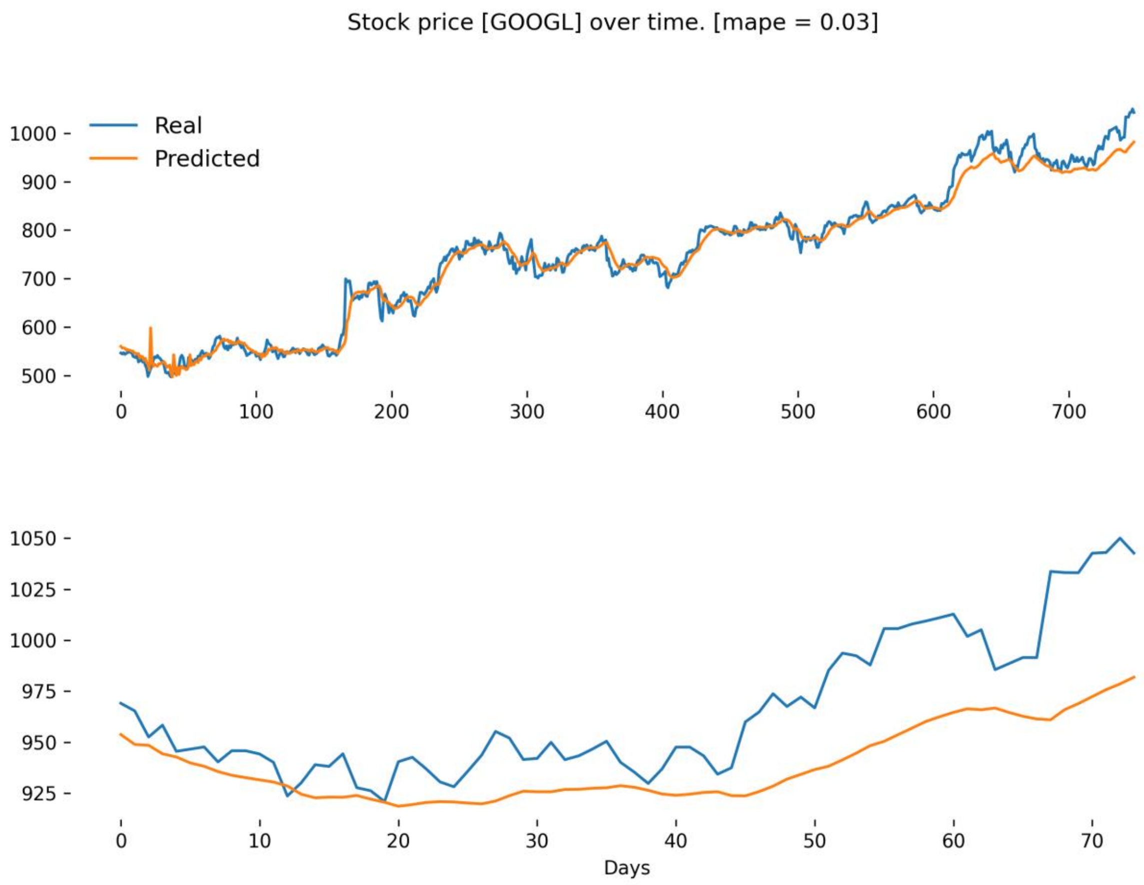
2 Does forecasting improve the portfolio?

Testing your model is still a must when doing forecasting!

Model Evaluation:

Model performance was measured using the mean absolute percentage error (MAPE).

A MAPE value of 0.03 implies that there is a deviation between actual and predicted values of 3%.

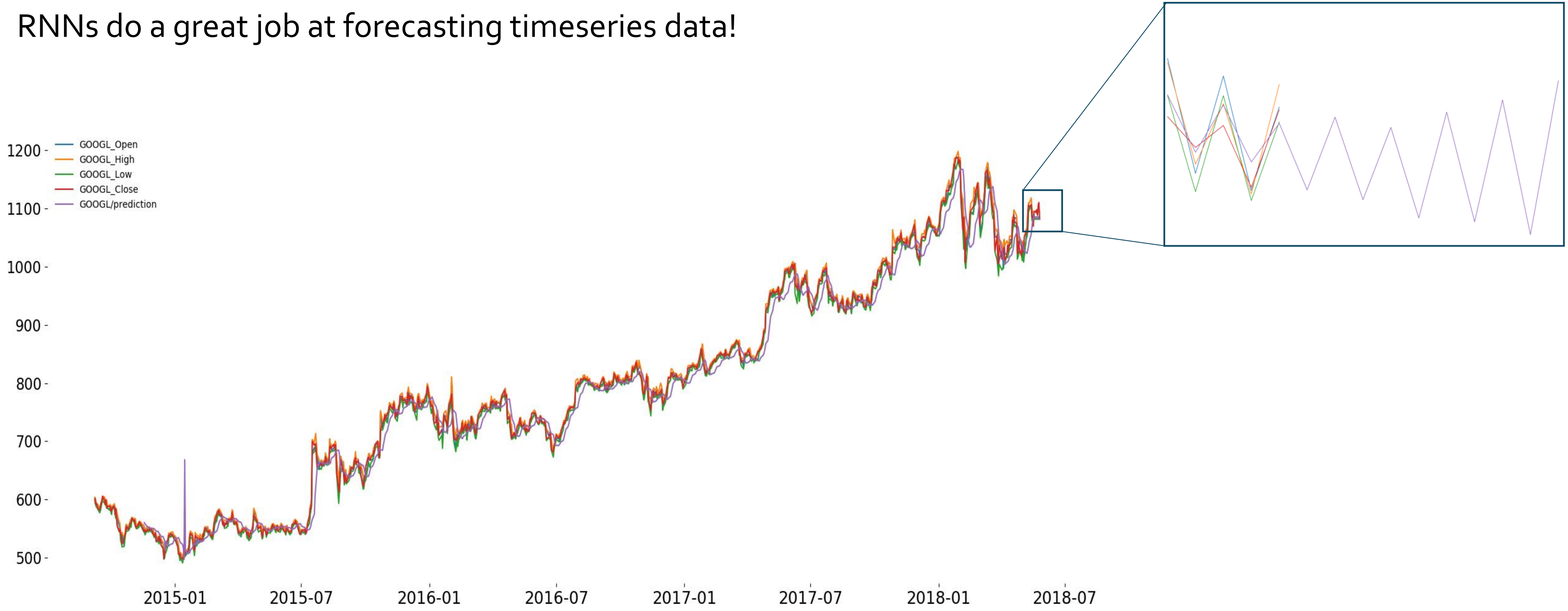


RNN model results fit on in-sample and out-of-sample dataset.

mVISE

② Does forecasting improve the portfolio?

RNNs do a great job at forecasting timeseries data!



RNN model results fit on entire dataset with 10-days out-of-sample forecast.

3 How to improve the risk calculation of a stock?

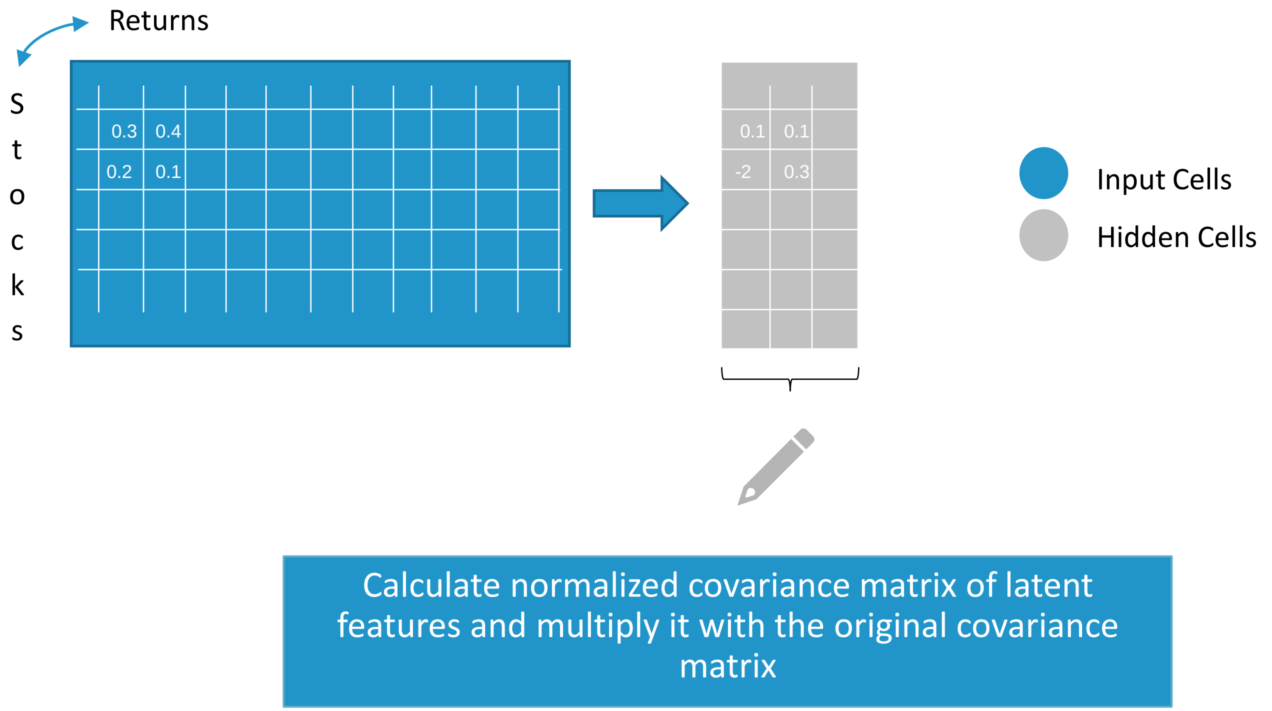
Latent features catch non-linearities and can be used to improve the sample covariance matrix!

- We transpose the input matrix and get a compressed time series in form of latent features.
- Calculating the normalized covariance of the latent feature vectors B, we are able to use this as a shrinkage estimator.

$$\hat{C} = B * C$$

Intuition:

Using the adjusted covariance matrix better captures non-linearities.

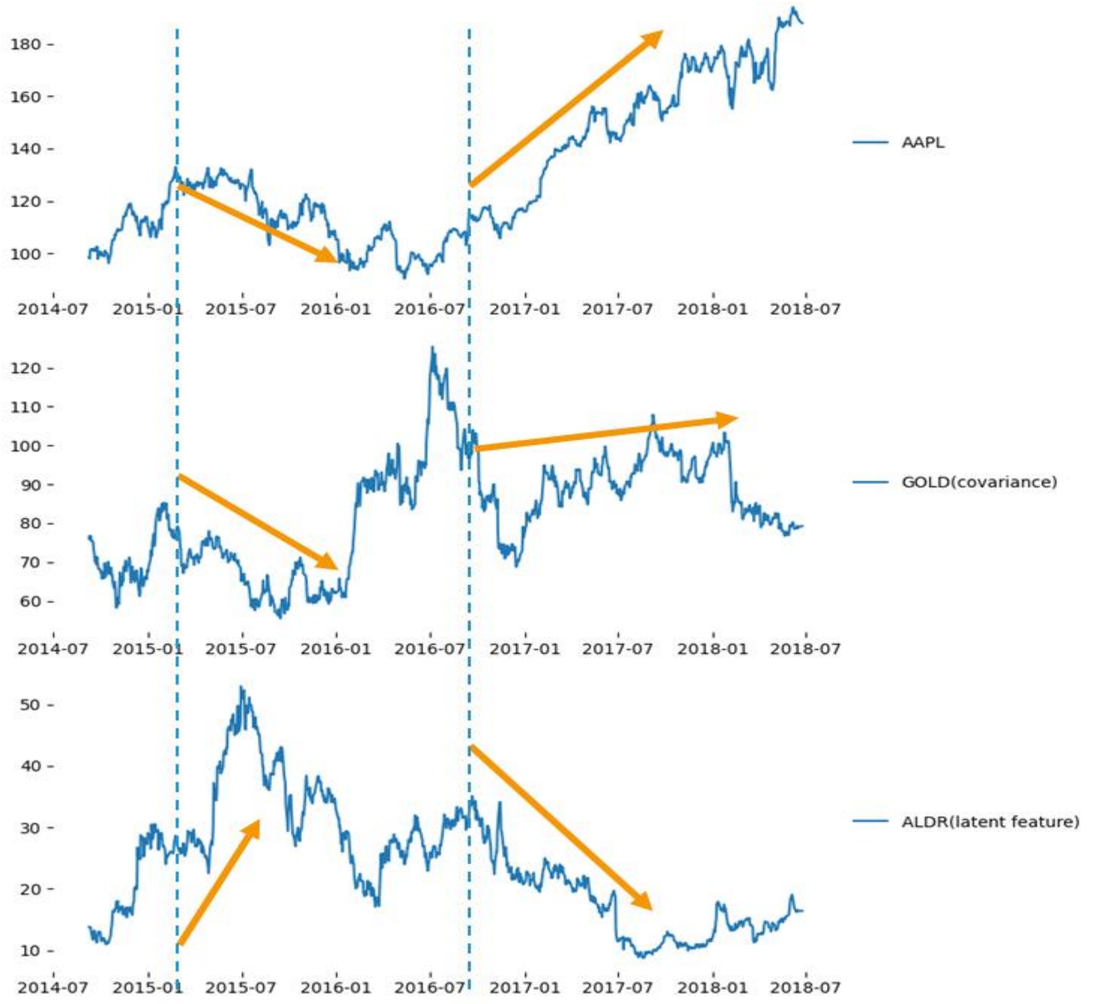


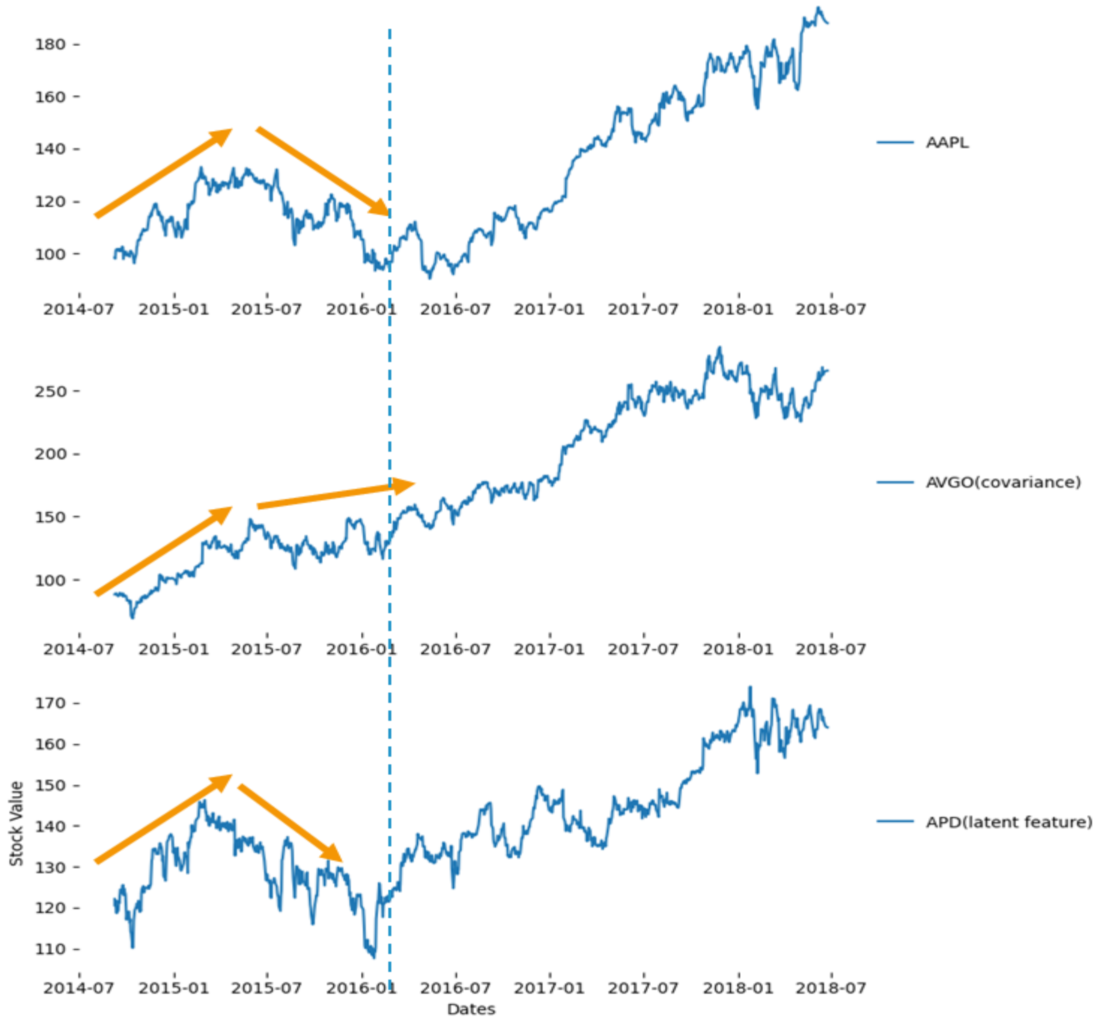
Autoencoder model with calculated covariance of latent features.

mVISE

3 How to improve the risk calculation of a stock?

Latent features catch non-linearities and can be used to improve the sample covariance matrix!

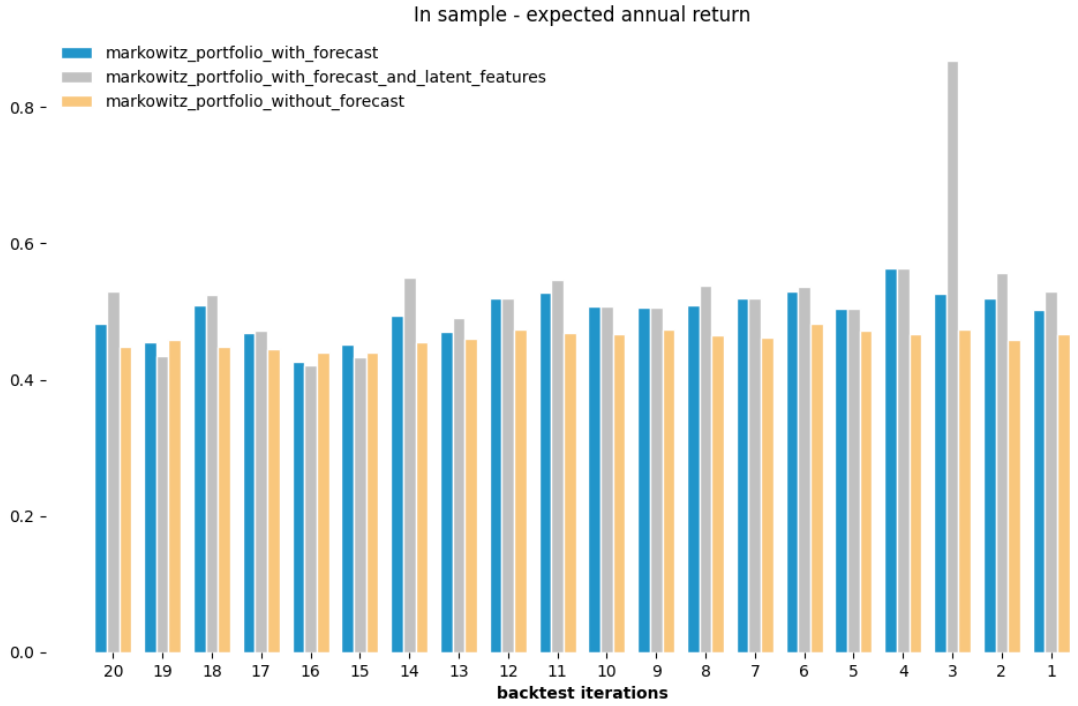


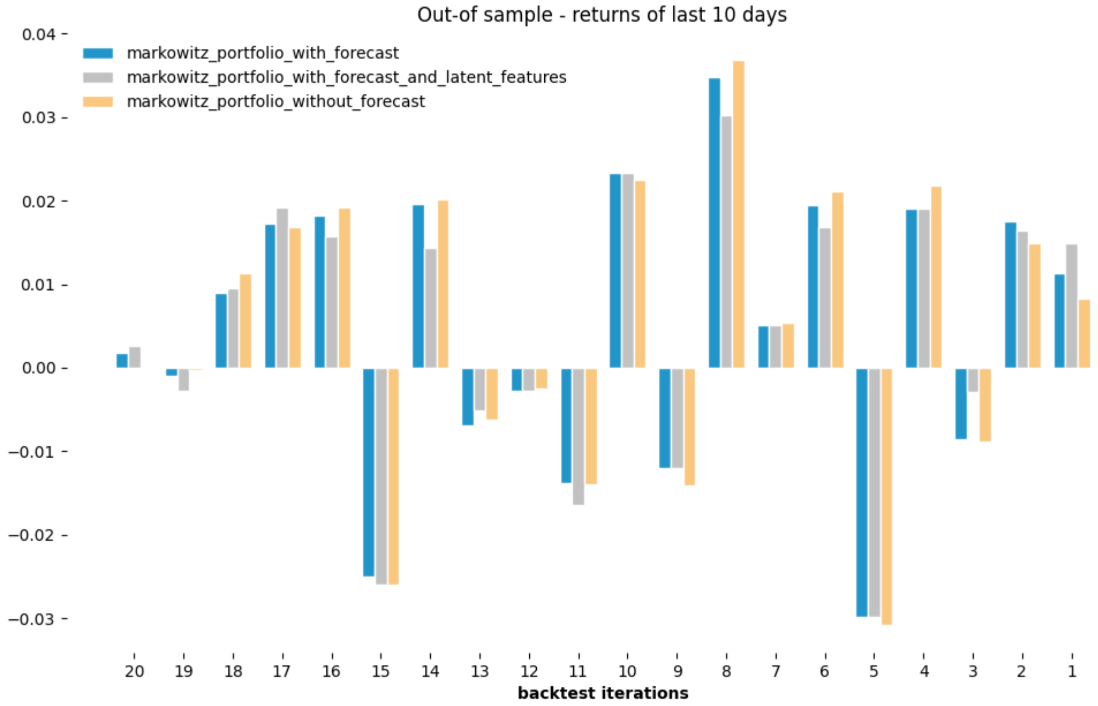


Baseline stock : APPL (Apple) compared to least (left) and most (right) related stocks

4 How to calculate an optimal portfolio?

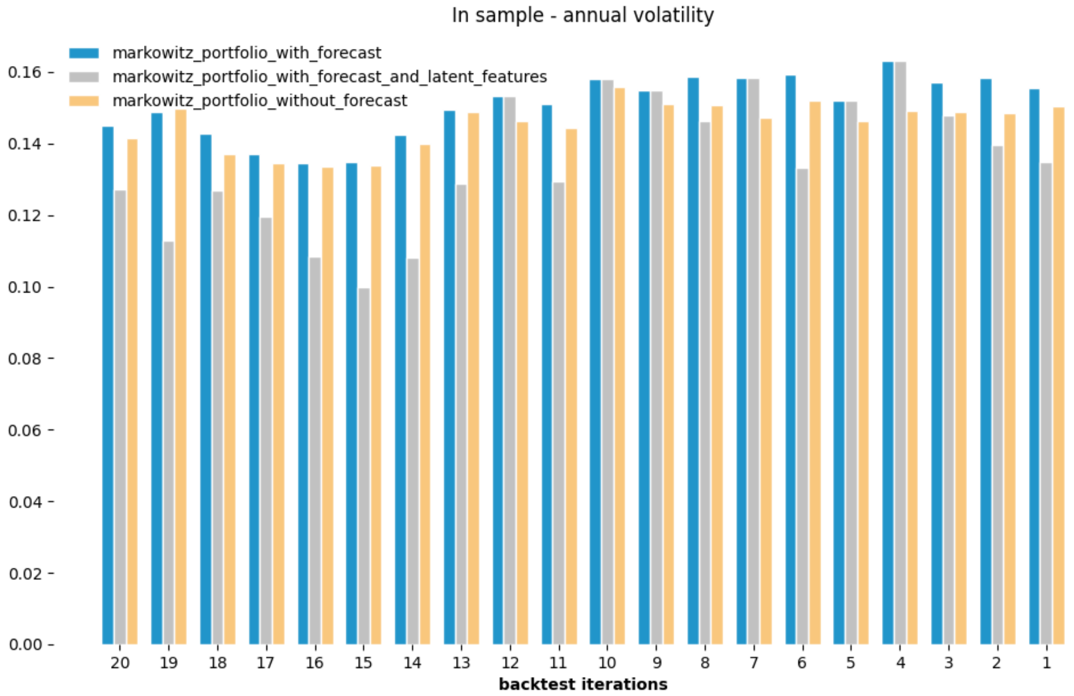
In-Sample results look good, out-of sample results do not look indicative.

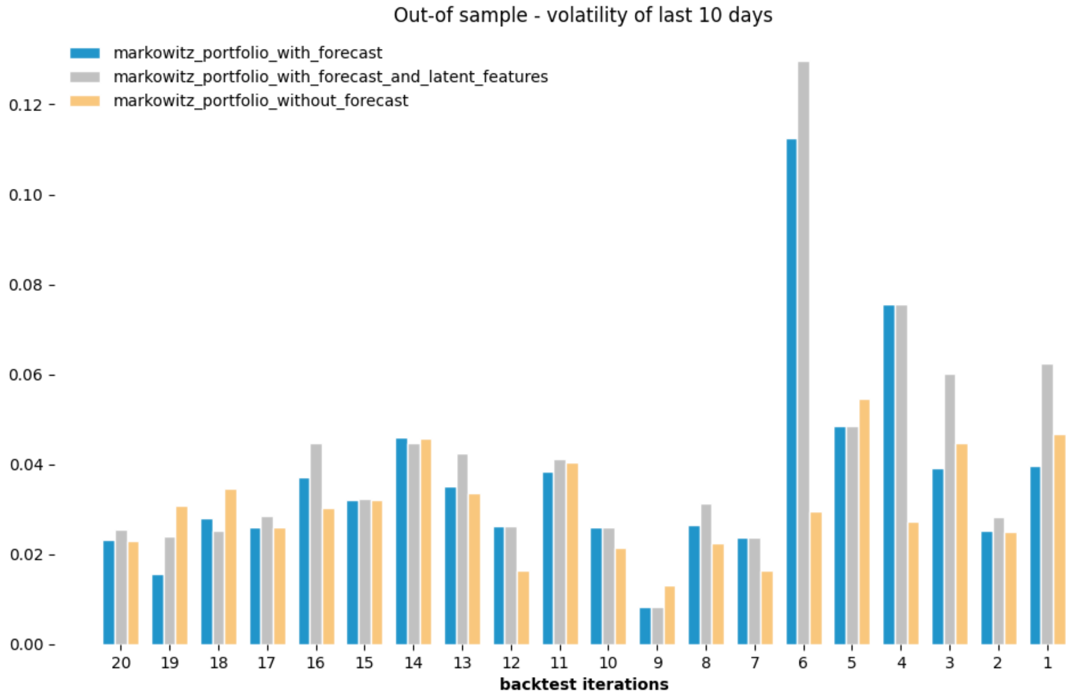




Expected annual return of Markowitz optimization using different input data (covariance matrix and returns)

4 How to calculate an optimal portfolio?

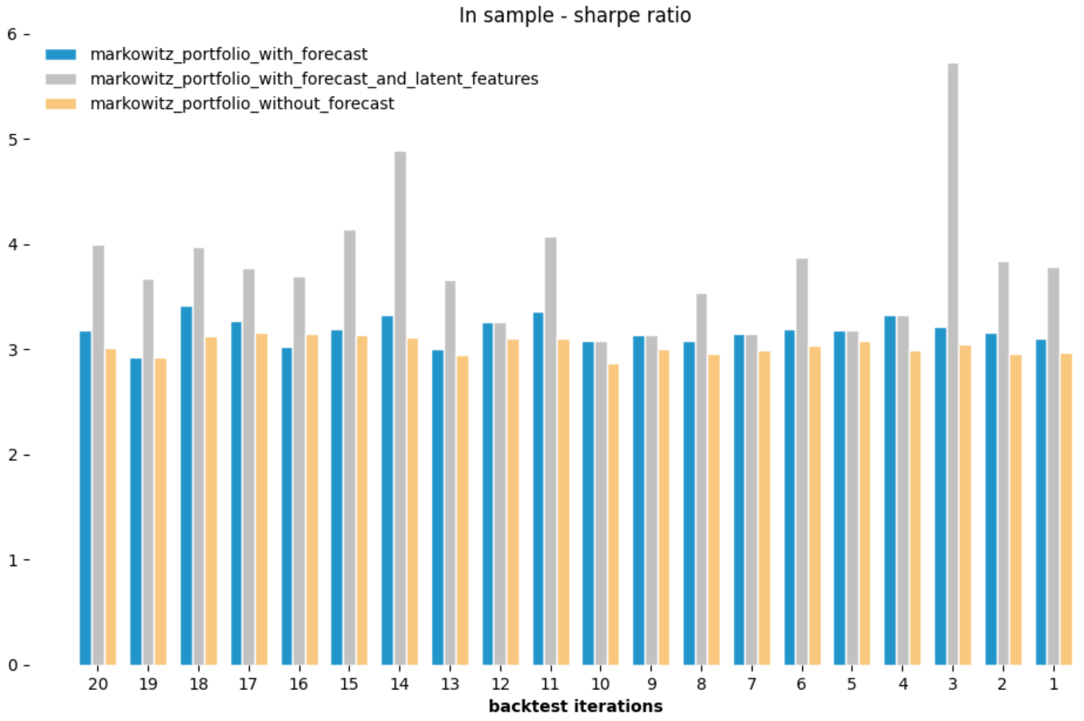


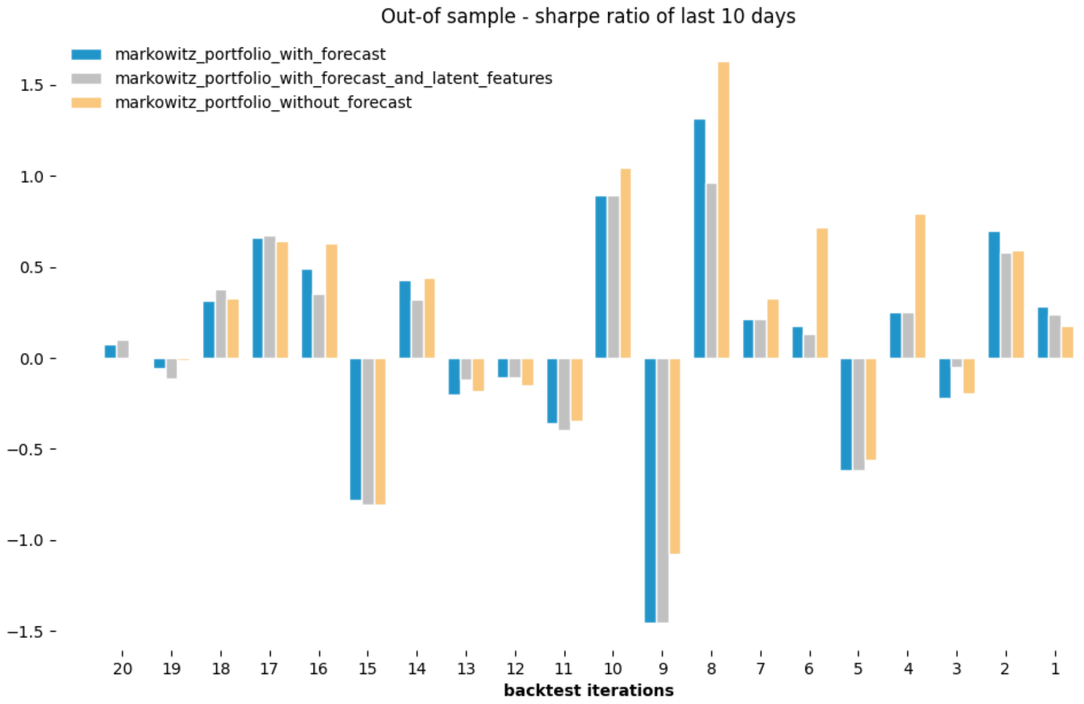


Annual volatility of Markowitz optimization using different input data (covariance matrix and returns)

mVISE

How to calculate an optimal portfolio?





Sharpe ratio of Markowitz optimization using different input data (covariance matrix and returns)

Conclusion and Future Research

Conclusion and Future Research

1 Which stocks to analyze?
→ Selecting stocks with the lowest reconstruction error **improves calculation** time and **shows better out-of-sample results**.

2 Does forecasting improve the portfolio?
→ Extending the dataset with a 10-day forecast leads to overall **higher portfolio results**.

3 How to improve the risk calculation of a stock?
→ Calculating the covariance of the latent features **reduces annual portfolio volatility** with similar or **increased stock returns**.

4 How to calculate an optimal portfolio?
→ The proposed model shows **superior results on the in-sample dataset**. The out-of-sample results may not be indicative.

References

- J. B. Heaton, N. G. Polson, & J. H. Witte. (2016). Deep Portfolio Theory.
- Werbos, P. (1990): Backpropagation Through Time: What It does and How to Do It. https://doi.org/10.1109/5.58337 Retrieved from http://axon.cs.byu.edu/~martinez/classes/678/Papers/Werbos_BPTT.pdf
- https://towardsdatascience.com/getting-rich-quick-with-machine-learning-and-stock-market-predictions-696802da94fe
- Olivier Ledoit & Michael Wolf, 2003. "**Honey, I shrunk the sample covariance matrix**," Economics Working Papers 691, Department of Economics and Business, Universitat Pompeu Fabra.

Q & A

1. Yes, my code is on github. https://github.com/QUER01/FinanceModule