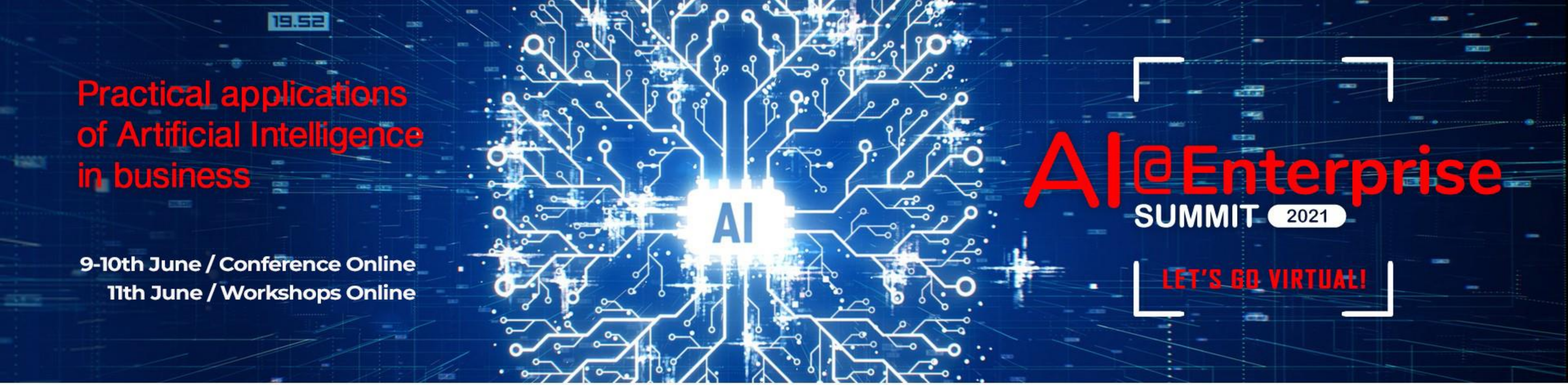


Stock Price Prediction and Portfolio Optimization Using Recurrent Neural Networks and Autoencoders

Julian Quernheim

Senior Consultant, BIVAL GmbH

Introduction

Goal:

- Apply deep learning to beat traditional portfolio optimization methods.

Problem Statements:

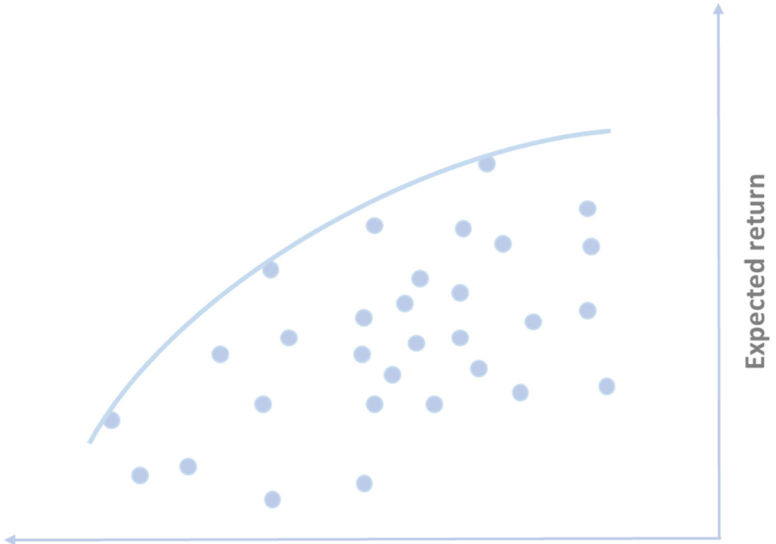
- Traditional methods for calculating an optimal stock portfolio **focus on historic data of stock returns.**
- Furthermore, risk is quantified **without capturing non-linearities** of the timeseries.

Requirements:

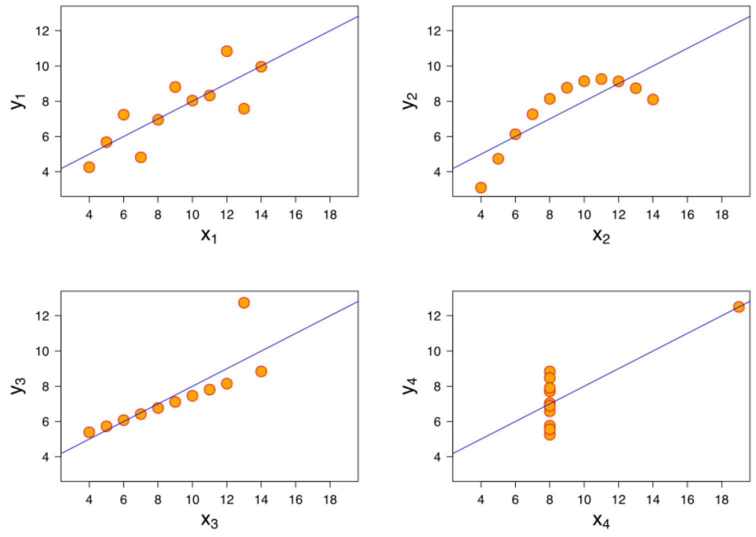
- A basic knowledge of deep learning and portfolio management is required.

Introduction

Investor can construct a portfolio of multiple assets that will **maximize returns** (r_i) for a given level of portfolio risk, **<u>but no future predictions are considered</u>**.



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**.



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

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

AI@Enterprise SUMMIT 2021 LET'S GO VIRTUAL!

Markowitz, 1952

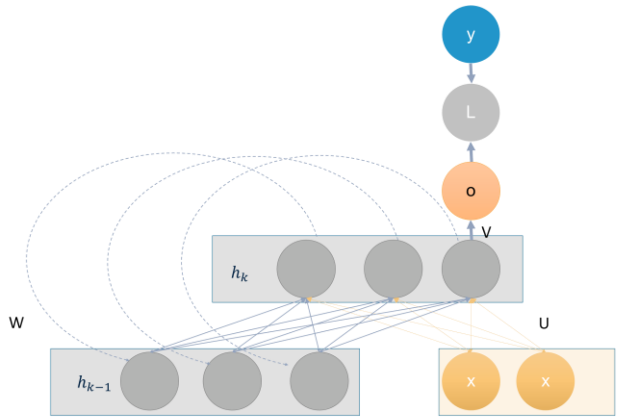
$$\text{minimize } C\, w^T w$$

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

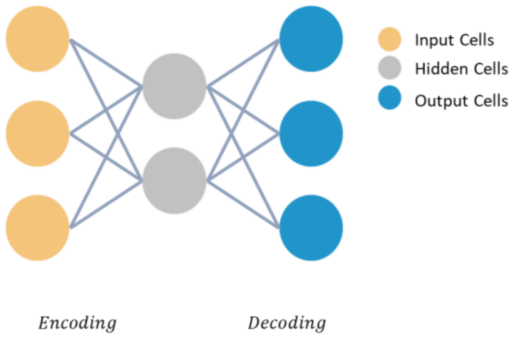
$$w^T \mathbf{1} = 1$$

$$w_i \geq 0$$

Recurrent neural networks



Autoencoders?



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

- Stock exchanges: NYSE and NASDAQ
- Tickers: 5685
- Range: 2014 -2018
- 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
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Transformed stock dataset

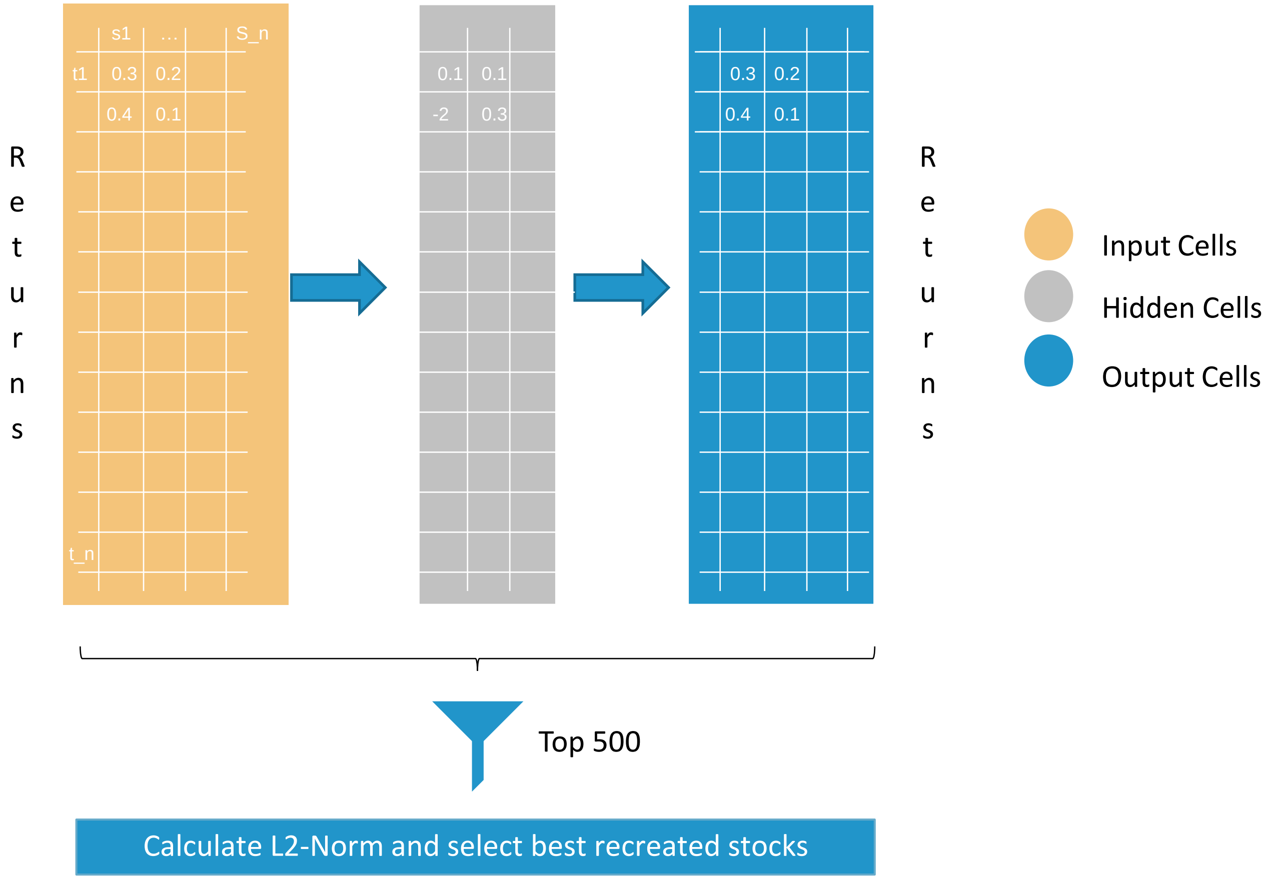
Original stock dataset

1 Which stocks to analyze?

Focus on stocks that move the market!

Intuition:

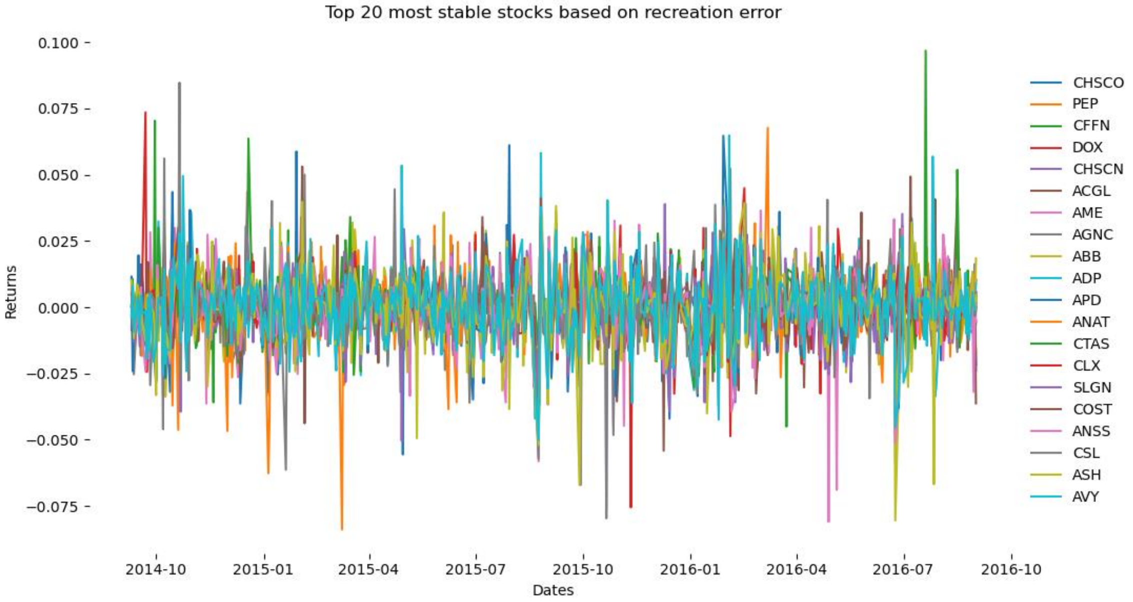
The stocks with the lowest recreation error (L2-norm) represent the market better. They are less volatile and are considered to be similar to large cap stocks.

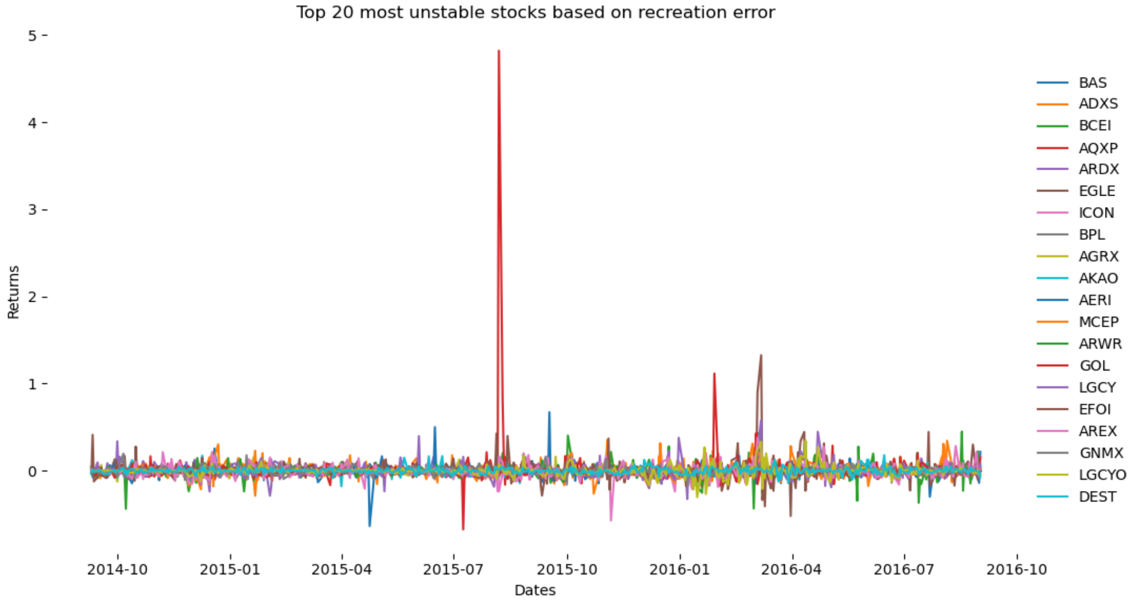


Autoencoder model with ranked recreation error.

1 Which stocks to analyze?

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

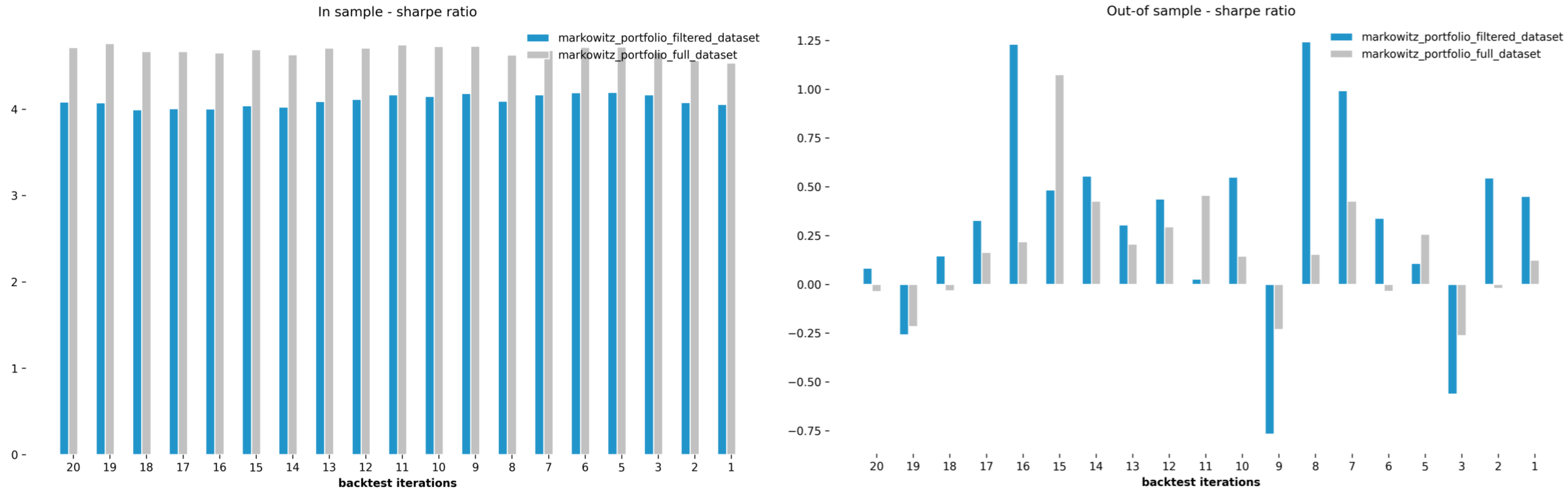




Top 20 most stable and most unstable stocks ranked by recreation error.

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.

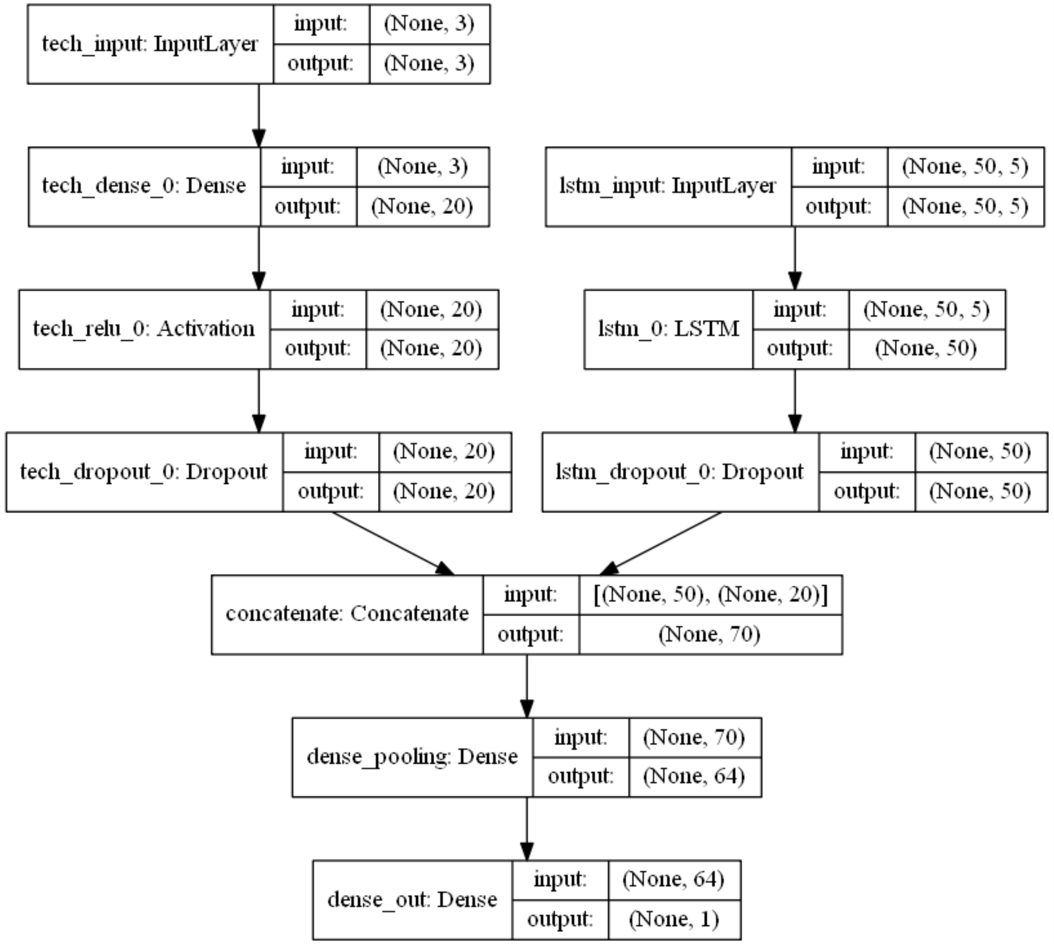
2 Does forecasting improve the portfolio?

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



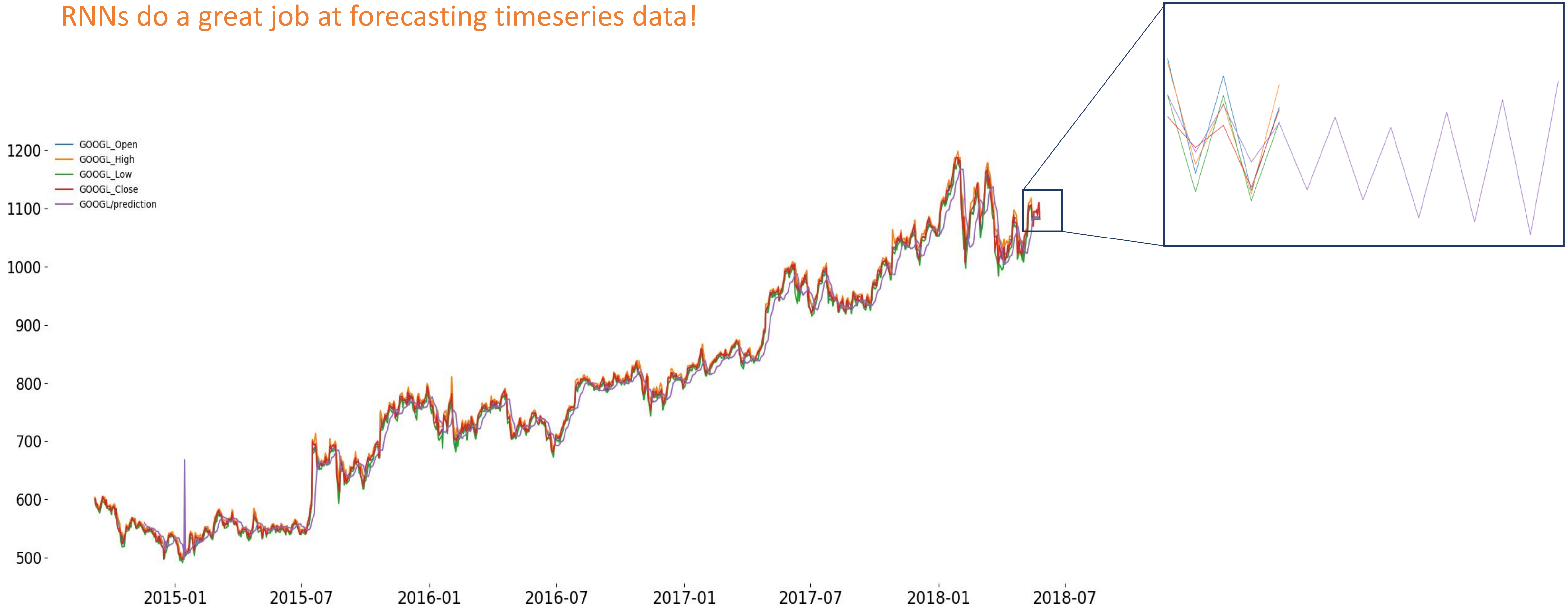
Keras RNN model.

Does forecasting improve the portfolio?

2

AI@Enterprise SUMMIT 2021 LET'S GO VIRTUAL!

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?

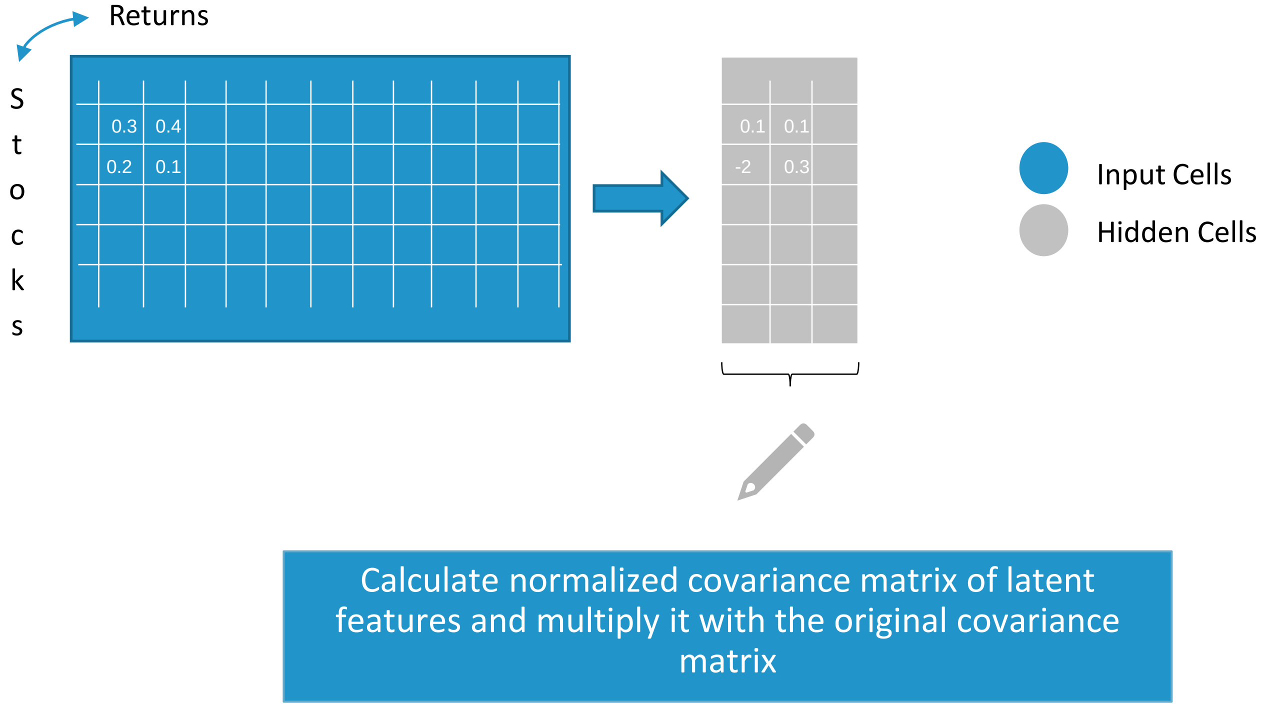
Latent features catch non-linearities and can be used to improve the sample covariance matrix!

- 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:

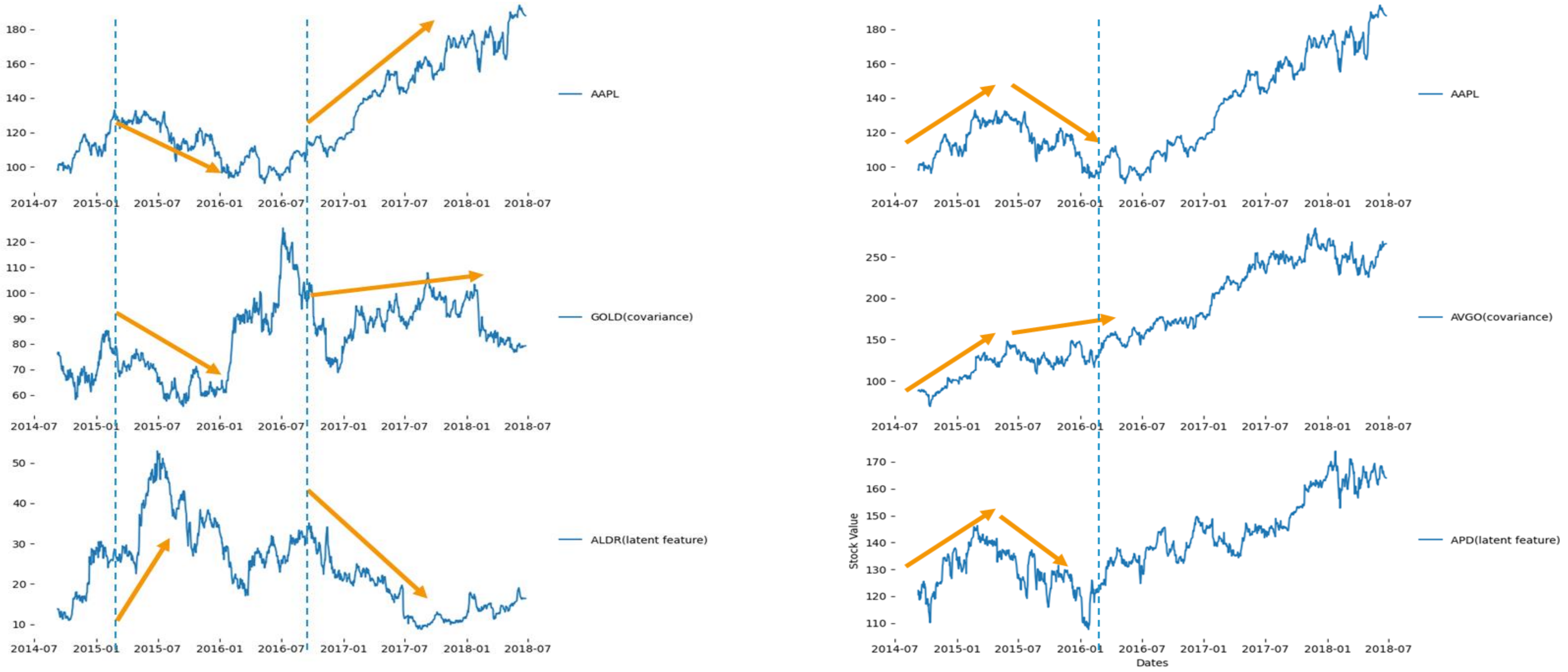
Using the adjusted covariance matrix better captures non-linearities.



Autoencoder model with calculated covariance of latent features.

3 How to improve the risk calculation of a stock?

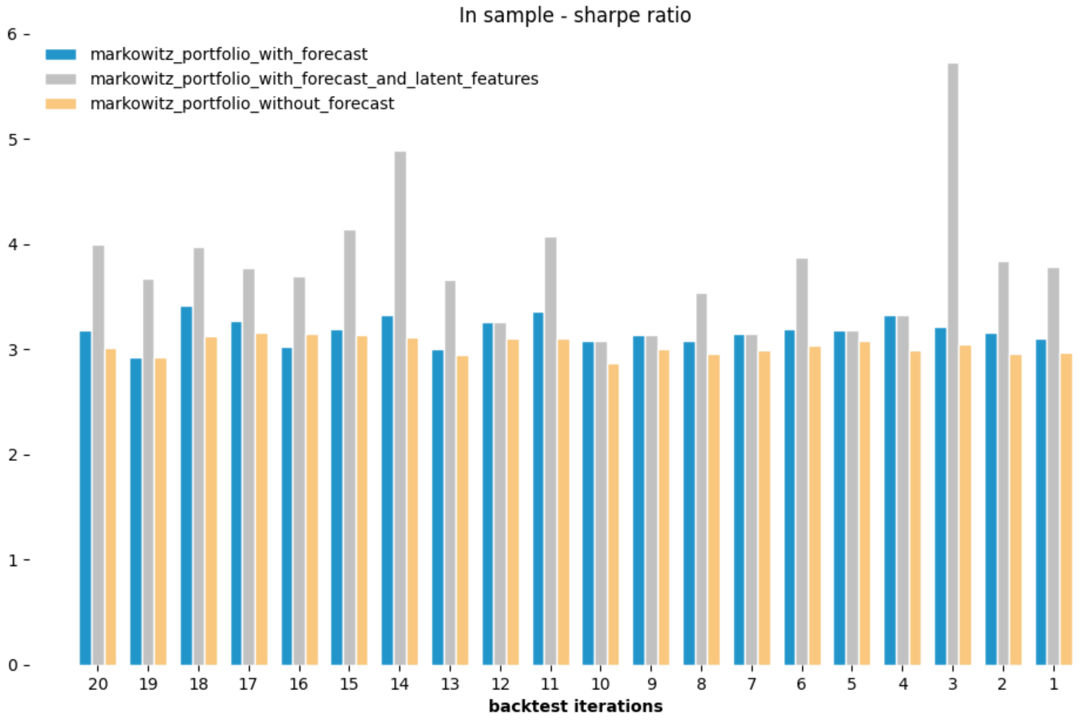
Latent features catch non-linearities and can be used to improve the sample covariance matrix!

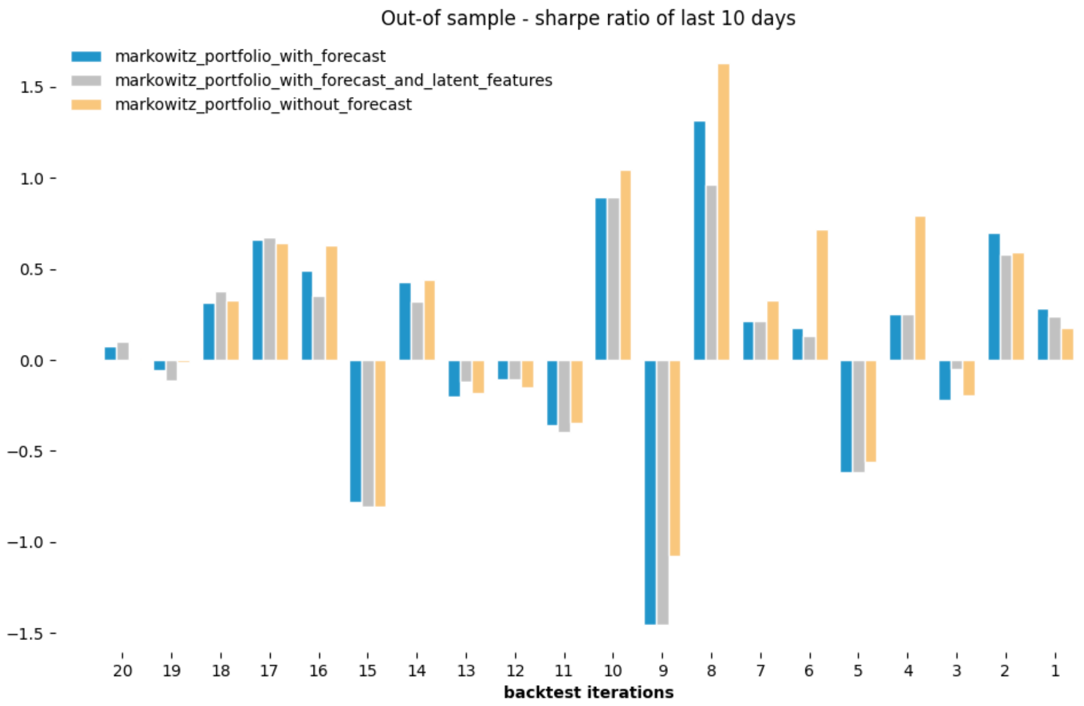


Baseline stock : APPL (Apple) compared to least (left) and most (right) related stocks

4 How to calculate an optimal portfolio?

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





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

Conclusion and Future Research

Conclusion and Future Research

AI@Enterprise SUMMIT 2021 LET'S GO VIRTUAL!

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

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