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

Predictive Analytics World 2020 Berlin

Julian Quernheim
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

- Markowitz (1952)
- Markowitz, Sharpe (CAPM) (1964)
- McKibbon and Rose (APT) (1973-1976)
- Black-Litterman (1991)
- Ledoit & Wolf (Shrinkage) (2002)
- Heaton et al. (Deep Portfolio) (2016)

Timeline of Portfolio Selection Methods
Literature Review

Timeline of deep learning methods

- Backpropagation (Werbos, 1974)
- RNN (Jordan, 1986)
- BBT (Werbos, 1990)
- LSTMs (Hochreiter, 1997)
- Heaton et al. (Deep Portfolio)
- Ledoit & Wolf, 2003 (Shrinkage)
- Autoencoder (Goodfellow et al., 2016)

Timeline of deep learning methods

Degree of applicability

- Markowitz
- Markowitz, Sharpe (CAPM)
- McKibbon and Rose (APT)
- 1952
- 1964
- 1973-1976
- 1991
- 2002
- 2016
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-Philosophy.5.htm
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.

Solutions

- Shrinkage Methods (Ledoit & Wolf, 2003)

\[ \hat{C} = \gamma C + (1 - \gamma) \times B \]

where:
- B is the shrinkage estimator and
- \( \gamma \) is the shrinkage constant

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

Modern portfolio theory (MPT) (Markowitz, 1952)

Linear programming formulation

\[
\begin{align*}
\text{minimize} & \quad Cw^T \quad w \\
\text{s.t.} & \quad w^T \mu \geq \mu_b \\
& \quad w^T 1 = 1 \\
& \quad w_i \geq 0
\end{align*}
\]

Where:

- the return of stock \( i \in \{1-n\} \) is defined as \( r_i \)
- the expected return vector is defined as \( \mu = \begin{pmatrix} E(r_1) \\ \vdots \\ E(r_n) \end{pmatrix} \), \( \mu_b \) 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 \( \text{cov}(r_1, r_2) \), and the covariance matrix of all stocks is \( n \) is \( \text{cov}(\mathbf{r}, \mathbf{r}) = \mathbf{C} \)

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

---

<table>
<thead>
<tr>
<th>Ticker</th>
<th>open</th>
<th>close</th>
<th>HL_diff</th>
<th>low</th>
<th>high</th>
<th>volume</th>
<th>date</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHH</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>B</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>C</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>D</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>E</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>F</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>G</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>H</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>I</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>J</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>K</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>L</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>M</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>N</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>O</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>P</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>Q</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>R</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>S</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>T</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>U</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>V</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>W</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>X</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>Y</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
<tr>
<td>Z</td>
<td>11.550000</td>
<td>11.580000</td>
<td>0.0315</td>
<td>11.530000</td>
<td>11.680000</td>
<td>48320000</td>
<td>2013-05-08</td>
</tr>
</tbody>
</table>

Original stock dataset
Dataset

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

Calculate L2-Norm and select best recreated stocks

Autoencoder model with ranked recreation error.
1. Which stocks to analyze?

Unstable stocks tend to be more volatile and have more unexpected spikes!
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.
Does forecasting improve the portfolio?

Data transformation is still part of the job!

50 day sliding window example for RNN model.
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?

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.
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.
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.

\[ \hat{C} = B \ast C \]

Intuition:
Using the adjusted covariance matrix better captures non-linearities.

Autoencoder model with calculated covariance of latent features.
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.
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)
How to calculate an optimal portfolio?

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


  Retrieved from http://axon.cs.byu.edu/~martinez/classes/678/Papers/Werbos_BPTT.pdf


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