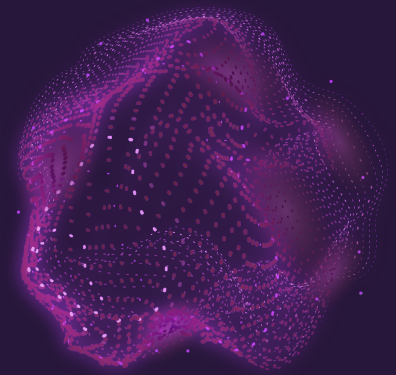
Advanced Machine learning (ML)



EigenPortfolio Pairs Trading

Ali Sadehinia, Bryan Rodas, Haining Fu

TABLE OF CONTENTS

History and Understanding pair trading

- **Origin:** Developed in the mid-1980s by quantitative analysts at Morgan Stanley.
- **Pioneers:** Gerry Bamberger and Nunzio Tartaglia's quantitative group.
- **Concept:** Identifying two securities whose prices move together historically. When the spread between their prices diverges, a trade is placed betting on the spread returning to its historical norm.
- **Evolution:** Originally used with stocks, but has expanded to other asset classes like ETFs, options, and futures.

- Mean Reversion
  - Mean reversion is the theory that prices and returns eventually move back towards the mean or average.

Pairs

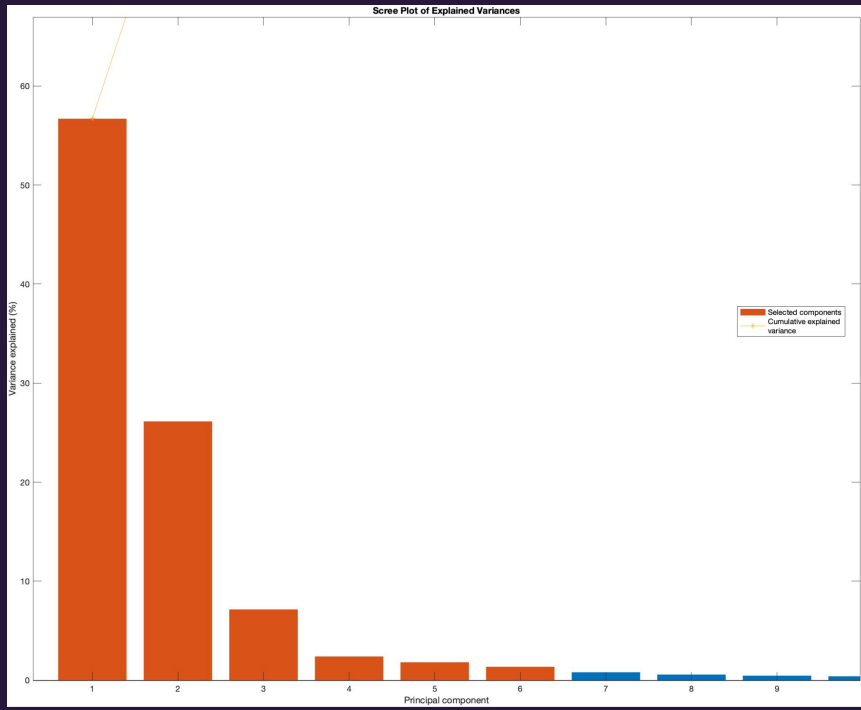
SPY ETF

Daily Closing Price
11-30-2020
To
11-30-2023

S&P 500 Companies

- Eigenportfolio of all the companies
- Companies that have not been added or removed in that time-frame
- 1st Principal Component as our portfolio

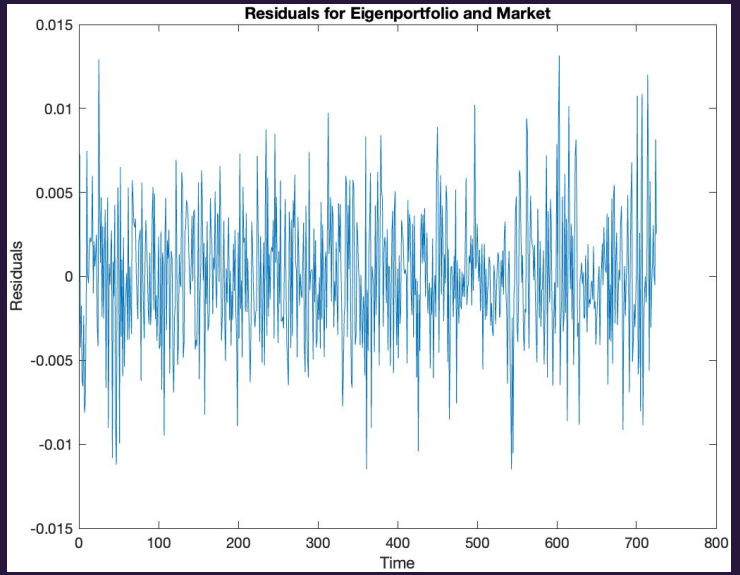
PCA Scree Plot (first 6 -> 95%)



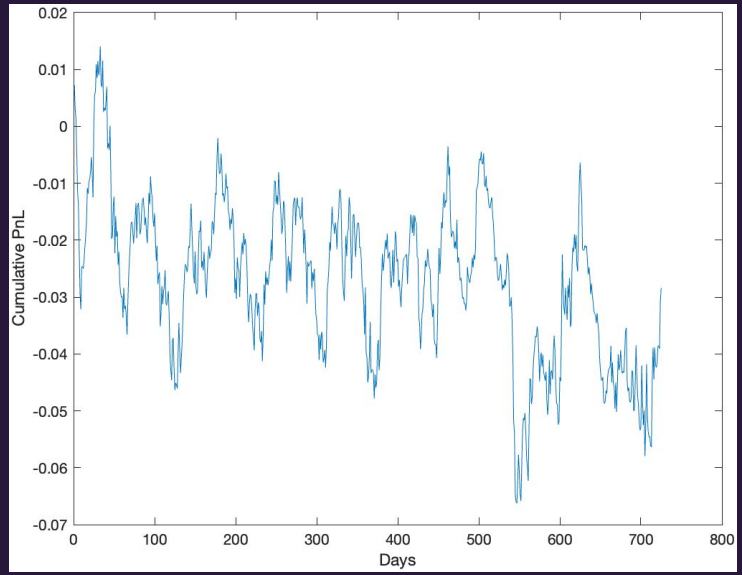
Methodology

1. PCA on the data
2. Use the 1st PC (~57% variance explanation)
3. Utilize linear regression on a sliding window of 30 and calculate the residuals ($1 eigenportfolio - $B SPY)
4. Calculate cumulative sum of the residuals and split the cumulative PnL into training and test
5. Train different LSTM models (GRU Layer here) on our standardized training data
6. Use the model to predict the future cumulative PnL
    a. Using our prediction for the last training data point to predict the future
    b. Using our test data to predict the future

Analysis and Findings



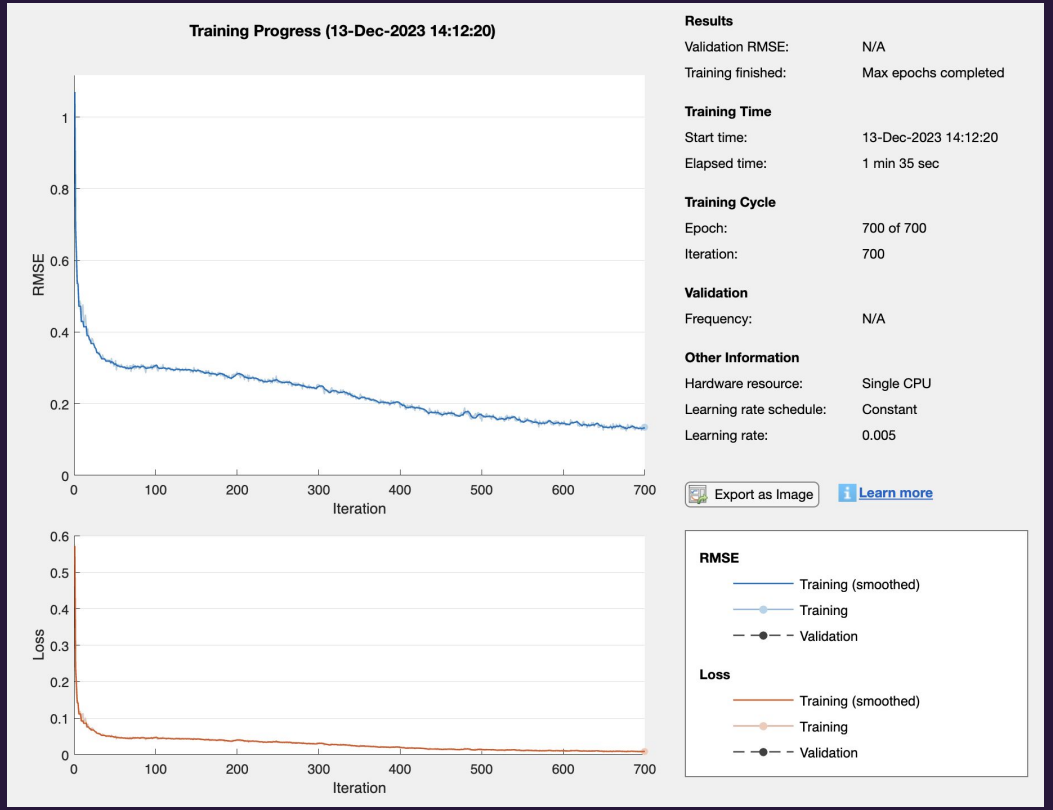
The residuals (our daily PnL for our strategy)



Cumulative sum of our residuals (our PnL on a rolling basis)

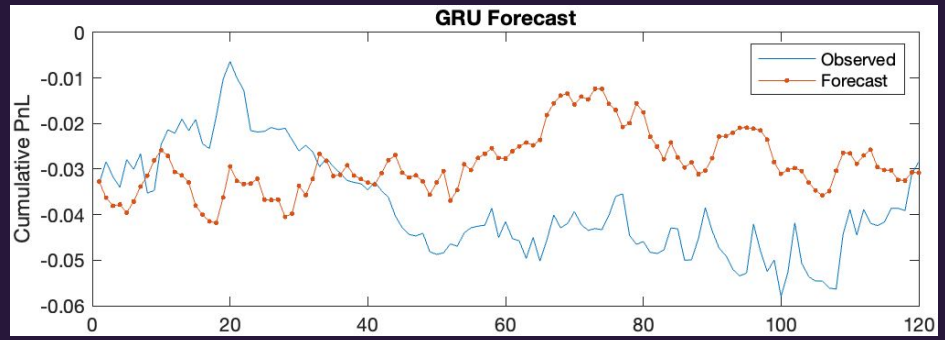
This is the RMSE over 700 iterations of 700 EPOCH with a GRU LSTM layer

The graph shows the process of minimizing the RMSE and loss function

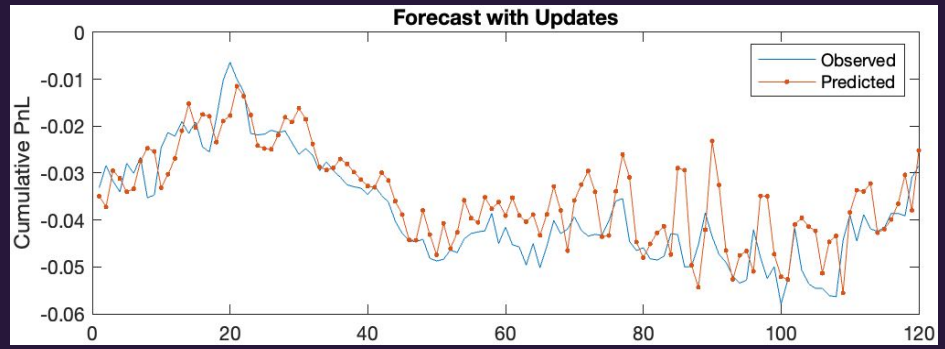


Analysis and Findings

Is this a bad strategy?
Our prediction based on our last training data point shows similar pattern in some sections between our forecast and observed data points



Is this a good model?
Our LSTM model's forecast has patterns that are following the observed data pretty well. RMSE of the model with our test data (0.143) is a value that is respectable.
We did not make the model too complicated since we are trying to stay away from overfitting.



Conclusion: Implications

- Other strategies/implementations to improve PnL:
  - Mean Reversion Strategy:
    - Signal: Trade when cointegration residuals significantly deviate from zero.
    - Logic: Buy underperforming and sell over performing stocks at certain residual thresholds.
  - Momentum Strategy:
    - Signal: Decisions based on historical returns.
    - Logic: Buy stocks with high past returns, sell those with low returns.
  - Hybrid Mean Reversion and Momentum:
    - Signal: Combine short-term mean reversion with long-term momentum signals.
    - Approach: Balance between exploiting short-term market inefficiencies and long-term trends.

Conclusion: Key Considerations:

- Risk Management: Incorporate stop-loss orders after entering the positions, position sizing, and other risk management techniques.
- Market Conditions: Be aware that different strategies may perform better in different market conditions.
- Transaction Costs: High-frequency strategies can incur significant costs, which must be factored into the strategy's profitability (especially during backtesting, although it would be inaccurate unless the trade is actually happening in the real world)

Thank you for your time!