Robust Multi-Period Portfolio Optimization of NASDAQ Stocks
Final Report

Theodoti Stoikou
Jan Philipp Girgott

Massachusetts Institute of Technology
Sloan School of Management
Optimization Methods (15.095)

December 7th, 2023

MIT MANAGEMENT SLOAN SCHOOL

Table of Contents

1 Problem Description

In our project, we solve an extension of the portfolio optimization problem presented in Recitation 7. The approach presented in that recitation considers only a one-period optimization of the portfolio. In reality, however, investors optimize their portfolios over multiple periods depending on the predicted returns. In our project, we created an optimal portfolio consisting of the top 500 NASDAQ stocks (ranked by market capitalization) with an initial investment of $100,000. The time horizon considered is the last year (Nov 2022 - Oct 2023), while the investor can transact daily.

2 Data

2.1 Downloading of Data from Yahoo! Finance

For our purposes, we downloaded the actuals for all NASDAQ-traded securities from Yahoo Finance using their Python API. The dataset contains the variables "Open" (opening price), "High" (maximum price of the day), "Low" (minimum price of the day), "Close" (closing price of the day), "Adj Close" (adjusted closing price of the day), "Volume" (number of shares traded on this day), and "company_name" (company name) for each of the approx. 12,000 securities. We preprocessed the dataset by (1) removing all ETFs, (2) deleting all stocks with less than two years of data, and (3) filtering for the top 500 stocks. We limited ourselves to the top 500 stocks to reduce the runtime for the models and to enhance interpretability.

2.2 Generation of Predictions using LSTM

Since it was challenging to retrieve detailed past forecasts which match our requirements from Bloomberg or the Refinitiv Workspace, we decided to generate our own predictions. We did this by leveraging the Long Short-Term Memory (LSTM) method to generate custom forecasts based on the historical data downloaded from Yahoo Finance. LSTM provides us with predictions which we require to implement the robust convex optimization formulation described later in this report.

3 Methodology

3.1 Lower Baseline: NASDAQ Index Fund Portfolio

As a lower baseline, we are considering the investment in a NASDAQ index fund. This means that the returns of the investor will correspond to the movements of the NASDAQ. We consider this approach as a lower baseline as it does not involve any active portfolio optimization but still achieves a very high degree of risk diversification.

3.2 Upper Baseline: Deterministic Model With Full Information

The upper baseline is a deterministic linear model that has perfect ex-post information, i.e., knows all closing prices. The full mathematical formulation can be found in the Appendix 1. Below, we provide a conceptual description and justification of the objective and each of the constraints.

The model is designed to maximize the end value of a financial portfolio over a predetermined number of time periods. This value is derived from the sum of the values of all stocks owned at that time and any cash remaining on hand at the final period (T).

At the outset of the investment period (period 1), the amount of each stock owned is solely determined by the initial purchase, symbolized by the first constraint. This is crucial as it establishes the initial position of the portfolio. To ensure that the initial stock purchase is within the investor's budget, constraint 2 restricts the first purchase. The total cost of purchasing stocks, including a value-based commission represented by α, should not exceed the initial cash available (C^0).

As the model progresses through subsequent periods, the amount of each stock owned is updated to reflect new purchases. The portfolio's stock composition in any period t is the sum of the amount owned at the end of the previous period and the amount bought in the current period (constraint 3). In particular, it prohibits short selling by ensuring that the total amount of any stock at any time is non-negative (constraint 4). Concurrently, the cash balance is adjusted to reflect these purchases. After each period, the cash on hand is recalculated as the remaining cash from the previous period minus the cost of current purchases and associated commission fees (constraints 5 and 6).

An integral aspect of the model is its risk management features. The model introduces a control mechanism on the transaction amount in any given period. It does this by limiting the total transaction amount to a fraction (δ) of the portfolio's total value, a strategy that mitigates risk by controlling the magnitude of investment or divestment actions within a single period (constraint 7). Diversification, a key tenet of risk management, is also enforced in the model. The value of any single stock in the portfolio at any given time is capped at a certain percentage (γ) of the total portfolio value. This constraint prevents over-concentration in any single asset, thereby reducing exposure to idiosyncratic risks associated with individual stocks.

We are not considering a constraint which models the impact of large-scale investors as we are unable to influence capital markets significantly with an initial endowment of only $100,000.

3.3 Robust Optimization Model Using Prediction Data

Our main model differs from the upper baseline model in three ways: (1) it uses prediction data for the daily stock prices instead of actuals, (2) it includes the variance to account for uncertainty, and (3) it includes robustness on the daily stock prices so that feasibility is retained even if the stock prices differ from the ones predicted. Please find the full formulation in Appendix 2.

In a real-life scenario, an investor does not have access to the actual future stock prices. However, they have access to analyst forecasts, a wide range of online tools, financial reporting documents, past stock prices. From these different pieces of information, investors can come up with stock price predictions which are close enough to the actual stock prices. To model this reality, we used a Long Short-Term Memory (LSTM) model to get predictions for stock prices. LSTM is a recurrent neural network widely used in Deep Learning which is very well suited for sequence prediction tasks. As visible in Appendix 3 the predictions are neither too far away nor to close to the actual prices.

The fact that the prices used are only predictions introduces uncertainty in the decisions of the investor considered in the model. A first way to reflect this uncertainty is to cap the maximum portfolio variance by a parameter σ^2. In practice, investors choose stocks with complementary covariance so that the overall portfolio variance remains low. To account for the fact that the value of our portfolio changes over time, we are not upper bounding the portfolio variance itself, but the portfolio variance relative to the portfolio value. This relationship is modeled by the first constraint of the robust optimization formulation. As a matter of fact, the choice of the parameter σ^2 greatly influences the result achieved by the model. Therefore, we estimated a range for it from the decisions made by the upper baseline model based on the actual stock prices. By calculating the relative portfolio variance for every period t, we received a range of relative portfolio variances between 1,828.3 up to 77,806.5. It needs to be borne in mind that that the hypothetical investor in the upper baseline does not face any uncertainty. Therefore, we decided to choose value of $\sigma^2 = 10{,}000$ at the lower end of the beforementioned range.

Eventually, we also introduced robustness to our model. To guarantee convexity of our problem and for ease of implementation, we decided to employ a weighted scenario approach with an uncertainty set based on the Central Limit Theorem. We decided to construct 10 scenarios which introduce stock-wise perturbations of maximum two standard deviations of this given stock. This is why the stock price matrix $P_{i,t}$ now transforms to a tensor $P_{i,t,s}$ where s denotes the respective scenario. Introducing robustness increases the number of constraints by a factor of approx. 10.

4 Results

We ran both models with the following parameter choices.

Initial investment:	C_0 = 100,000
Commission per dollar:	α = 0.0001
Maximum percentage of individual stock:	γ = 0.2
Limit on transaction amount:	δ = 0.05
Maximum relative portfolio variance:	σ^2 = 10,000

For this parametrization, the upper baseline model achieves a final portfolio value of $2.8m whereas the robust model increases the portfolio value to $730k. Over the same time horizon, the lower baseline model increases the initial investment to $120k. The value achieved by the robust model, thus, corresponds to a sixfold increase in portfolio value compared to the lower baseline. In the following, we will compare the upper baseline and robust models regarding total portfolio value over time, portfolio diversification over time and portfolio composition in the last period.

4.1 Total Portfolio Value Over Time

The line graph in Appendix 4 depicts the progression of total portfolio value across three different investment models over one year. The robust model, represented by the red line, shows a consistent yet moderate increase in wealth, maintaining a steady trajectory between the upper and lower baselines. Despite not reaching the heights of the black line (upper baseline), which indicates a more aggressive growth pattern, the robust model outperforms the lower baseline, which remains relatively static and close to the initial investment level. The upper baseline's significant growth suggests an idealized strategy with perfect hindsight, while the robust model's steadier curve implies a more realistic and risk-managed approach, which nonetheless succeeds in achieving growth above the minimal gains of the lower baseline. The robust model's steadiness suggests that it has likely incorporated mechanisms to buffer against market downturns, showing resilience in times of volatility, as evidenced by its relative stability compared to the upper baseline's more pronounced fluctuations.

4.2 Portfolio Diversification Over Time

In line graph in Appendix 5, we see a comparison of the portfolio diversification between two investment strategies over a year. The upper baseline holds a lower number of stocks, indicating a more concentrated investment strategy with the number of stocks owned staying within the 10 to 15 range. This suggests a targeted approach, potentially focusing on a select group of stocks that are expected to perform exceptionally well regardless of risk (as the upper baseline does not face any risk). On the other hand, the black line, representing the robust model, demonstrates a more diversified portfolio approach with the number of stocks owned fluctuating between 30 to over 40. This spread of investments could signify a strategy designed to mitigate individual stock volatility and sector-specific risks, aiming for a steadier performance through diversification. The robust model's approach aligns with the principle of not putting all eggs in one basket, whereas the upper baseline's strategy might be banking on the strong performance of a curated selection of stocks.

4.3 Portfolio Composition in the Last Period

The pie charts in Appendix 6 compare the portfolio composition between the robust model and the upper baseline at the end of the last period. The robust model's portfolio is more diversified, with a more equitable distribution of asset allocation, where the largest holding doesn't exceed 14.3%. In contrast, the upper baseline's portfolio is more concentrated, with its largest holding at 20%. Thus, the constraint mandating that an individual shareholding may only take a maximum of 20% is tight in this case. Such diversification in the robust model could be a strategic choice to minimize unsystematic risk. The robust model's spread across various sectors could reflect a strategy designed to capitalize on a broader range of growth opportunities, whereas the upper baseline seems to focus on fewer, potentially higher-performing stocks.

5 Discussion of Results

5.1 Managerial Implications and Recommendations

In general, the results of this report provide three main implications for portfolio managers.

Importance of Portfolio Diversification for Risk Averse Investors with Incomplete Information
The model results confirm that portfolio diversification is one of the most effective ways to hedge against uncertainty for investors. In this context, the fundamental principle of diversification is not just about expanding the number of stocks in a portfolio but about strategically selecting a mix that minimizes risk and maximizes returns. To optimally find this mix, it is important for portfolio managers to understand the risk appetite of their customers very well. Only in this way, it is possible to find a suitable value of σ^2 when speaking in optimization terms.

Robustness as an Additional Way to Hedge Against Imprecise Predictions
Next to controlling the portfolio variance, incorporating robustness emerges as a second very effective way of controlling uncertainty. In general, the model decisions are only as good as the input data it is provided. Acknowledging the limitations of forecasting in a complex market, robustness hedges against the risks of imprecise predictions rather than optimizing for a specific predicted outcome. If quantitative portfolio managers are able to model uncertainty sets which correctly reflect the volatility of the stocks considered, robustness can provide an edge over other investors.

Importance of Effective Liquidity Management/Divesting at the Right Time
Effective liquidity management and timely divesting involves recognizing the right time to exit certain positions – whether to realize gains, cut losses, or rebalance the portfolio in response to shifting market dynamics. Both the upper baseline model and the robust optimization model master these decisions (see drops in portfolio value in Appendix 4 and cash on hand in Appendices 7/8) and decide effectively when it is more beneficial to hold a larger cash position then to invest in stocks.

5.2 Limitations and Avenues for Further Exploration

We acknowledge that our model is only a simplification of reality and has a couple of limitations. First, both models assume that an investor makes all investment decisions based on the information that was available at $t = 0$. In fact, new information becomes available throughout the investment period which enables the investor to potentially make better decisions. Second, investors can usually also choose from other securities apart from stocks (e.g., ETFs, pension funds, real estate funds, bonds). Widening the types of securities available and potentially varying the model parameters can be exciting ways to get a more nuanced analysis of the dynamics at play. Last, one could try different methods of incorporating robustness in the model (e.g., norm uncertainty sets) and compare the outcomes with each other. In this context, it would be interesting to see how well the different approaches can hedge against varying levels of imprecise forecasts.

6 Appendix

6.1 Full Formulation of the Upper Baseline Model

Indices

$$
\begin{aligned}
i = 1, \ldots, n: &\quad \text{set of stocks} \\
t = 1, \ldots, T: &\quad \text{set of time periods}
\end{aligned}
$$

Decision Variables

$$
\begin{aligned}
x_{it}: &\quad \text{amount of stock } i \text{ bought during period } t \\
z_{it}: &\quad \text{amount of stock } i \text{ owned at the end of period } t \\
c_t: &\quad \text{cash on hand at the end of period } t
\end{aligned}
$$

Input Data and Parameters

$$
\begin{aligned}
P_{it}: &\quad \text{price of stock } i \text{ during period } t \\
C^0: &\quad \text{initial cash on hand for investment in stocks} \\
\alpha: &\quad \text{commission paid per dollar traded} \\
\gamma: &\quad \text{maximum percentage of individual stock of total value of portfolio} \\
\delta: &\quad \text{limit on transaction amount}
\end{aligned}
$$

Model

$$
\begin{aligned}
\max_{c,x,z} \quad & \sum_{i=1}^{n} P_{i,T} z_{i,T} + c_T \\
\text{s.t.} \quad & z_{i,1} = x_{i,1} && \forall i \in [1,n] \\
& (1+\alpha) \sum_{i=1}^{n} P_{i,1} x_{i,1} \leq C^0 \\
& z_{i,t} = z_{i,t-1} + x_{i,t} && \forall i \in [1,n], t \in [2,T] \\
& z_{i,t-1} + x_{i,t} \geq 0 && \forall i \in [1,n], t \in [2,T] \\
& c_1 = C^0 - \sum_{i=1}^{n} P_{i,1} x_{i,1} - \alpha \sum_{i=1}^{n} P_{i,1} |x_{i,1}| \\
& c_t = c_{t-1} - \sum_{i=1}^{n} P_{i,t} x_{i,t} - \alpha \sum_{i=1}^{n} P_{i,t} |x_{i,t}| && t \in [2,T] \\
& -\delta(c_t + \sum_{i=1}^{n} z_{i,t-1} P_{i,t}) \leq \sum_{i=1}^{n} |x_{i,t}| P_{i,t} \leq \delta(c_t + \sum_{i=1}^{n} z_{i,t-1} P_{i,t}) && t \in [2,T] \\
& P_{i,t} z_{i,t} \leq \gamma \sum_{i=1}^{n} (P_{i,t} z_{i,t}) && \forall i \in [1,n], t \in [2,T] \\
& c_t, z_{i,t} \geq 0 && \forall i \in [1,n], t \in [1,T]
\end{aligned}
$$

6.2 Full Formulation of the Robust Model

Indices

$$
\begin{aligned}
i = 1, \ldots, n: &\quad \text{set of stocks} \\
t = 1, \ldots, T: &\quad \text{set of time periods} \\
s = 1, \ldots, S: &\quad \text{set of scenarios of stock prices}
\end{aligned}
$$

Decision Variables

$$
\begin{aligned}
x_{it}: &\quad \text{amount of stock } i \text{ bought during period } t \\
z_{it}: &\quad \text{amount of stock } i \text{ owned at the end of period } t \\
c_t: &\quad \text{cash on hand at the end of period } t
\end{aligned}
$$

Input Data and Parameters

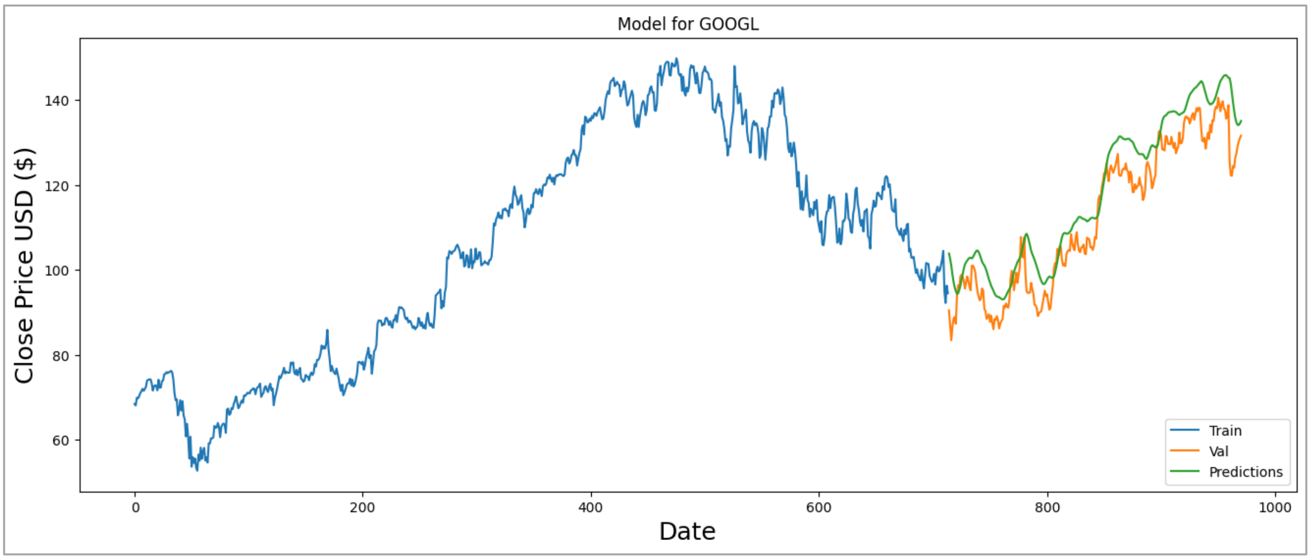
$$
\begin{aligned}
P_{its}: &\quad \text{price of stock } i \text{ in scenario } s \text{ during period } t \\
C^0: &\quad \text{initial cash on hand for investment in stocks} \\
\Sigma_{ijs}: &\quad \text{covariance matrix of stocks } (i,j) \text{ in scenario } s \\
\alpha: &\quad \text{commission paid per dollar traded} \\
\gamma: &\quad \text{maximum percentage of individual stock of total value of portfolio} \\
\delta: &\quad \text{limit on transaction amount} \\
\sigma^2: &\quad \text{maximum portfolio variance}
\end{aligned}
$$

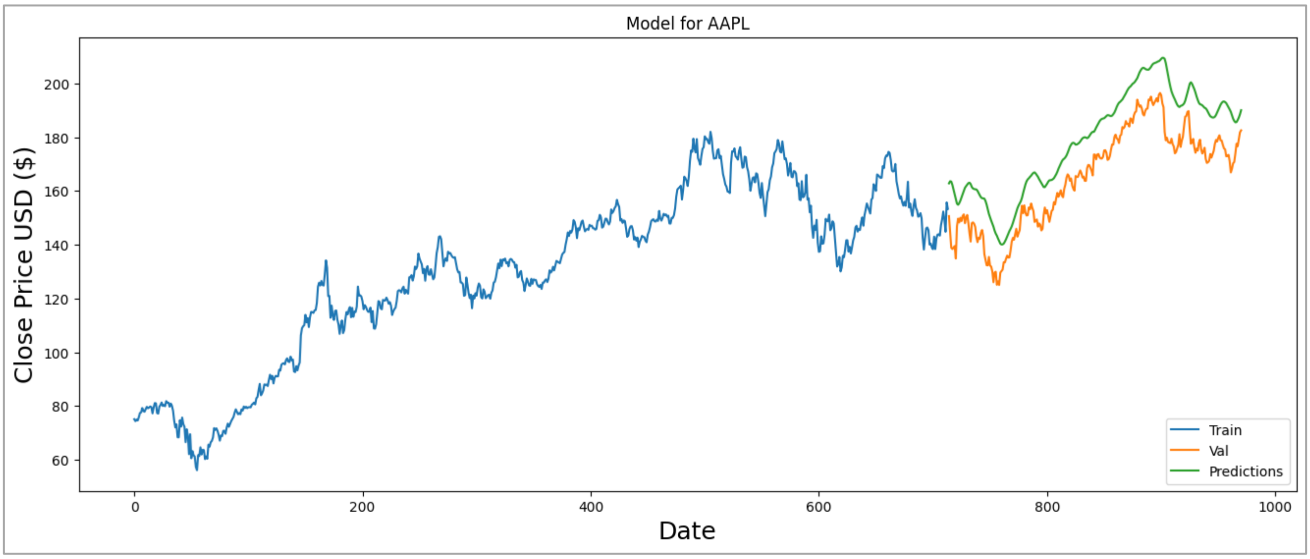
Model

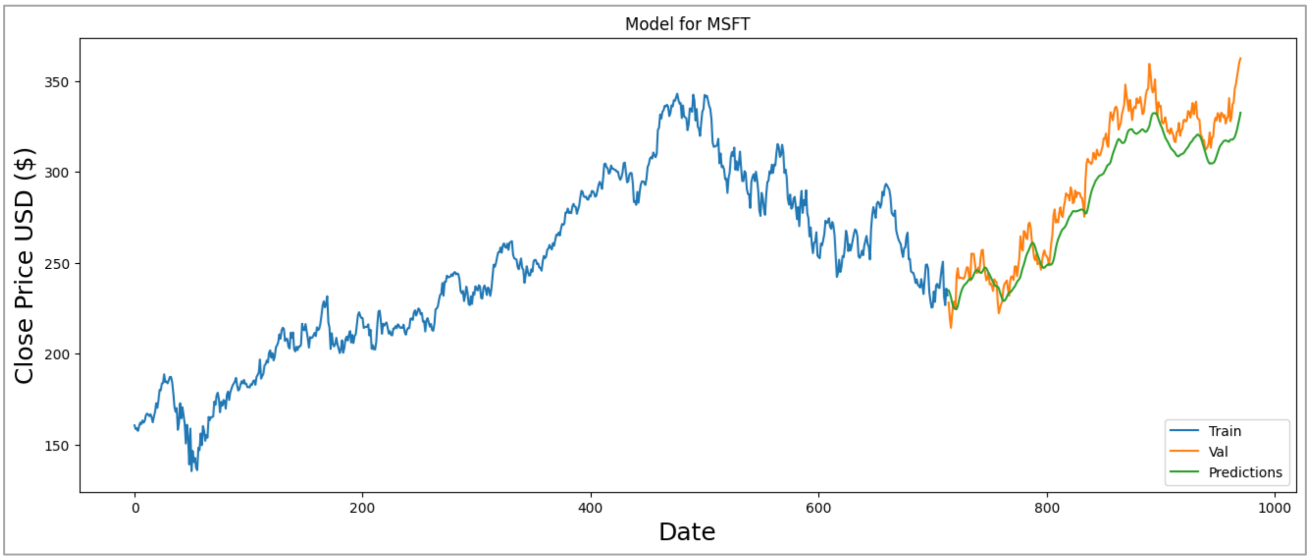
$$
\begin{aligned}
\max_{c,x,z} \quad & \frac{1}{S} \sum_{i=1}^{n} \sum_{s=1}^{S} (P_{i,T,s} z_{i,T}) + c_T \\
\text{s.t.} \quad & \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} z_{i,t} \Sigma_{i,j,s} z_{j,t}}{\sum_{i=1}^{n} z_{i,t} P_{i,t,s}} \leq \sigma^2 && \forall t \in [1,T], s \in [1,S] \\
& z_{i,1} = x_{i,1} && \forall i \in [1,n] \\
& (1+\alpha) \sum_{i=1}^{n} (P_{i,1,s} x_{i,1}) \leq C^0, && \forall s \in [1,S] \\
& z_{i,t} = z_{i,t-1} + x_{i,t} && \forall i \in [1,n], t \in [2,T] \\
& z_{i,t-1} + x_{i,t} \geq 0 && \forall i \in [1,n], t \in [2,T] \\
& c_1 = C^0 - \sum_{i=1}^{n} (P_{i,1,s} x_{i,1}) - \alpha \sum_{i=1}^{n} (P_{i,1,s} |x_{i,1}|) && \forall s \in [1,S] \\
& c_t = c_{t-1} - \sum_{i=1}^{n} (P_{i,t,s} x_{i,t}) - \alpha \sum_{i=1}^{n} (P_{i,t,s} |x_{i,t}|) && t \in [2,T], s \in [1,S] \\
& -\delta(c_t + \sum_{i=1}^{n} z_{i,t-1} P_{i,t,s}) \leq \sum_{i=1}^{n} (|x_{i,t}| P_{i,t,s}) \leq \delta(c_t + \sum_{i=1}^{n} z_{i,t-1} P_{i,t,s}) && t \in [2,T], s \in [1,S] \\
& P_{i,t,s} z_{i,t} \leq \gamma \sum_{i=1}^{n} (P_{i,t,s} z_{i,t}) && \forall i \in [1,n], t \in [2,T], s \in [1,S] \\
& c_t, z_{i,t} \geq 0 && \forall i \in [1,n], t \in [1,T]
\end{aligned}
$$

MIT MANAGEMENT SLOAN SCHOOL

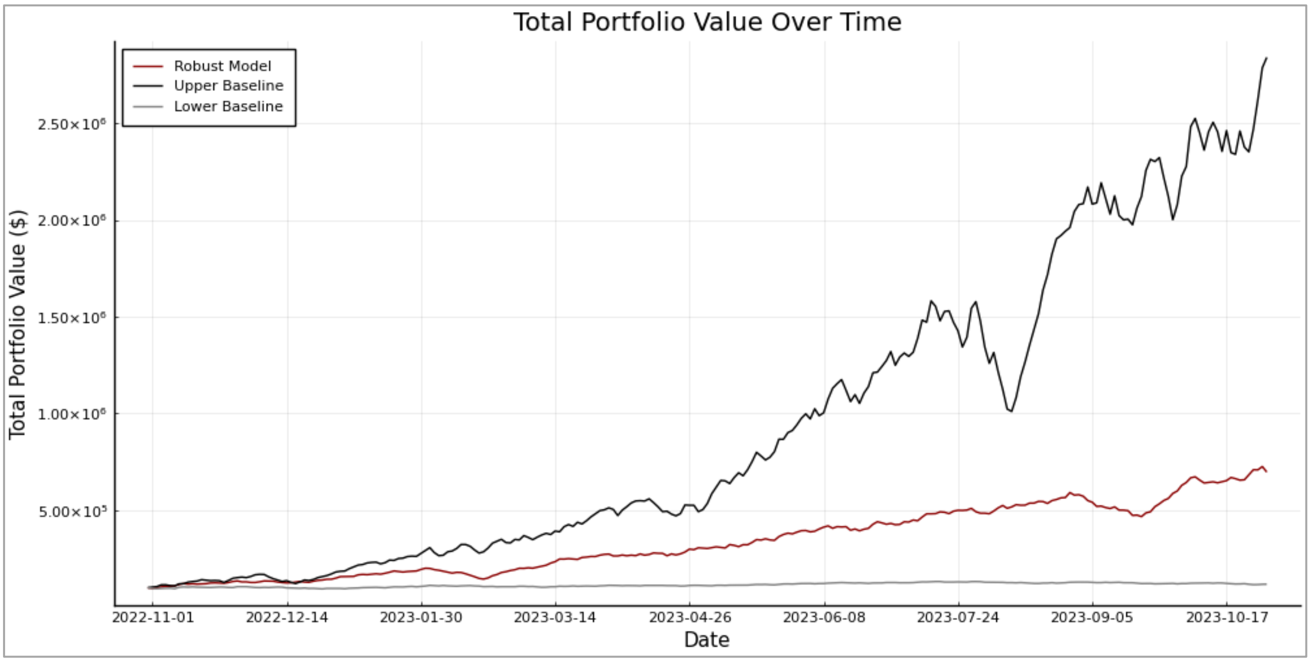
6.3 Example Output of the LSTM Model



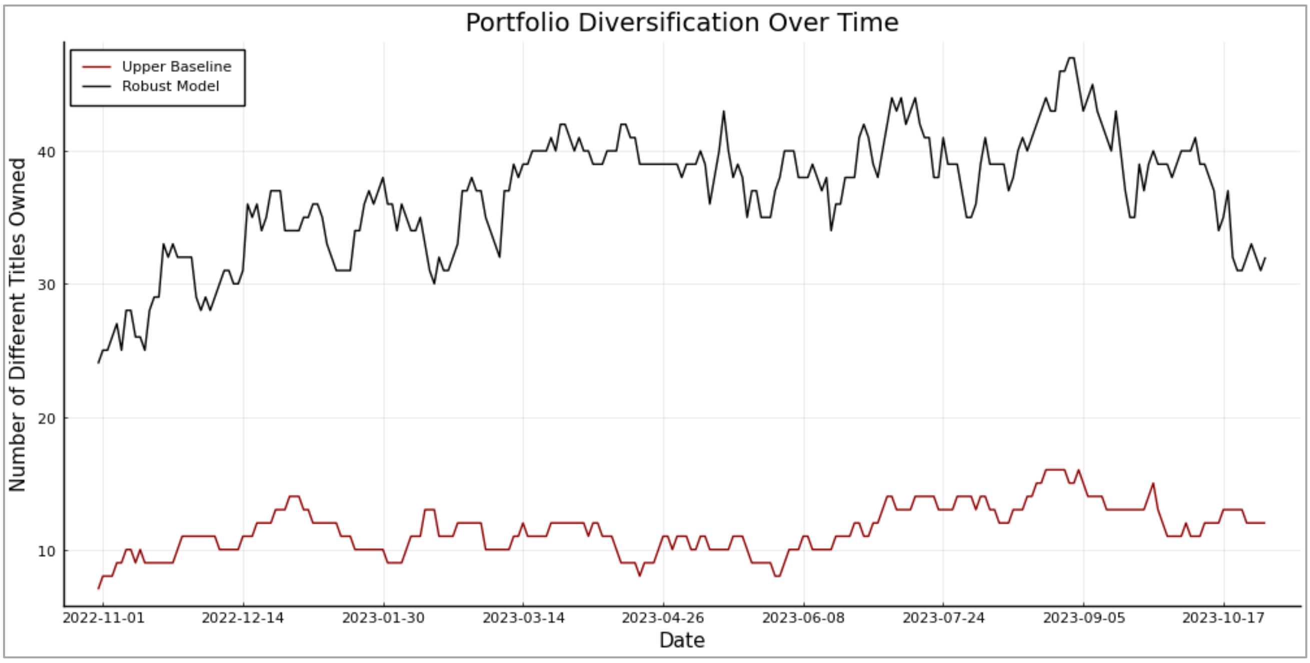




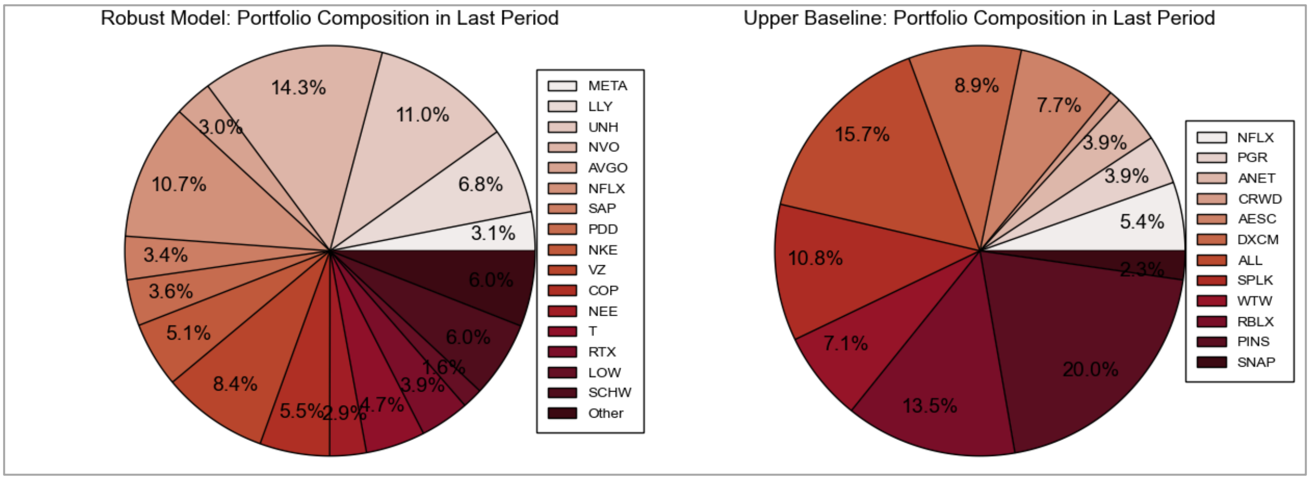
6.4 Total Portfolio Value Over Time



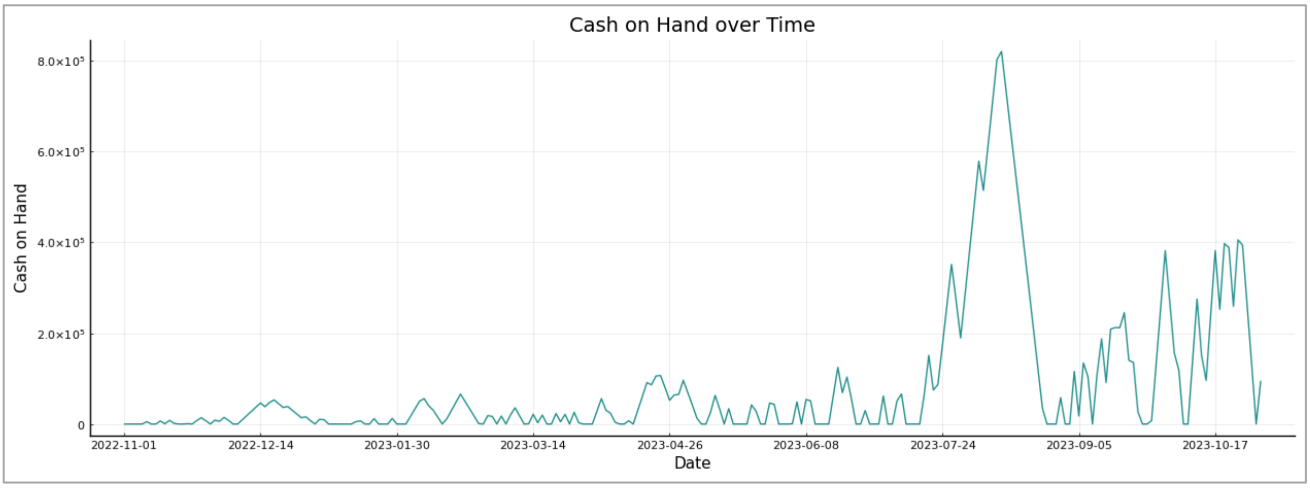
6.5 Portfolio Diversification Over Time (Number of Different Titles Owned)



6.6 Portfolio Composition in Last Period



6.7 Cash on Hand (Baseline Model)



6.8 Cash On Hand (Robust Model)