Simulating Competition in Trading

Mario Hernandez

August 2024

Abstract

1 Introduction

The problem of selecting a profitable subset of stocks from a given set S is a fundamental challenge in financial analysis and investment strategy. The goal is to develop a function $f(S)$ that identifies such a subset based on certain criteria. This problem is particularly relevant in the context of stock price distributions exhibiting positive skewness, where the potential for significant positive returns exists.

Consider a set of stocks $S = \{S_0, S_1, \ldots, S_n\}$. We aim to find a subset $S' \subseteq S$ such that the stocks in S' have a high probability of achieving profitable returns. This problem can be examined under two scenarios: when the time horizon T_f is constrained and when it is unconstrained.

1.1 Unconstrained Time Horizon

In the scenario where the time horizon T_f is not constrained, the stock prices can be modeled as stochastic processes evolving indefinitely. Let S_t represent the stock price at time t. Given a positive skew in the stock price distribution, the expected return $E(S_t)$ over an unbounded time horizon T can be described by the integral:

$$E(S_t) = \int_{-\infty}^{\infty} s \cdot f(s)\, ds$$

where $f(s)$ is the probability density function of the stock price.

As T tends to infinity, the probability that the stock price will achieve at least a certain percentage change δ approaches 1:

$$\lim_{T \to \infty} P\left(\frac{S_T}{S_0} \geq 1 + \delta\right) = 1$$

This is a direct consequence of the law of large numbers and the cumulative nature of returns over time.

1.2 Constrained Time Horizon

When the time horizon T_f is constrained, we need to consider the finite period within which the stock must achieve the desired return δ. The expected return over a constrained time horizon T_f is given by:

$$E(S_{T_f}) = \int_{-\infty}^{\infty} s \cdot f_{T_f}(s)\, ds$$

where $f_{T_f}(s)$ is the probability density function of the stock price at T_f.

The probability of the stock reaching the percentage change δ within the constrained time T_f is:

$$P\left(\frac{S_{T_f}}{S_0} \geq 1 + \delta\right)$$

As T_f increases, this probability tends to 1, although it may not be 1 for finite T_f. The rate of convergence depends on the skewness and volatility of the stock price distribution.

2 Graph Representation of Stock Prices

Formally, let $S = \{S_0, S_1, \ldots, S_n\}$ be a set of stocks. For each stock S_i, we construct a graph $G_i = (V_i, E_i)$, where:

- V_i is the set of nodes representing time intervals.
- E_i is the set of edges representing the temporal sequence of these intervals.
- Each node $v \in V_i$ contains attributes (H_v, L_v, C_v, O_v), corresponding to the high, low, close, and open prices, respectively.

2.1 Temporal Sequence

The nodes V_i are ordered by time, creating a directed edge from v_t to v_{t+1} for each t. This structure captures the chronological progression of stock prices.

2.2 Multi-Stock Integration

To integrate multiple stocks, we can visualize each stock's graph as a string of price history along the time dimension. By adding more stocks, we form a two-dimensional plane where:

- The x-dimension represents time, shared across all stocks.
- The y-dimension represents the individual stock's price history.

Adding a third z-dimension can represent the highs and lows of the y-dimension's price history, resulting in a three-dimensional object (x, y, z).

3 Methodology

3.1 Probability Calculation

For each stock i, the model predicts the probability of hitting a positive return $P(y_i^{class} = 1)$. This probability is obtained from the softmax output of the classification head of the neural network:

$$P(y_i^{class} = 1) = softmax(f_{class}(x_i))_1$$

where $f_{class}(x_i)$ represents the logits produced by the classification head for stock i.

3.2 Expected Time

The expected time $E[T_i|y_i^{class} = 1]$ to hit the positive return is predicted by the regression head of the neural network:

$$E[T_i|y_i^{class} = 1] = f_{time}(x_i)$$

where $f_{time}(x_i)$ represents the predicted time output by the regression head for stock i.

3.3 Utility Function

The utility function U_i for each stock i is defined to balance the likelihood of hitting a positive return and the speed of achieving this return. The utility function is given by:

$$U_i = \frac{P(y_i^{class} = 1)}{E[T_i|y_i^{class} = 1] + \epsilon}$$

where ϵ is a small constant to prevent division by zero.

3.4 Stock Selection

To select the top-performing stocks, we rank the stocks based on their utility values. The selection method involves the following steps:

1. Compute the utility U_i for each stock i.
2. Sort the stocks in descending order of their utility values.
3. Select the top k stocks with the highest utility values.

Formally, the subset S of the top k stocks is given by:

$$S = \arg \max_{S' \subseteq \{1,\ldots,N\}} \sum_{i \in S'} U_i$$

where N is the total number of stocks, and k is the number of stocks to be selected.

4 Background

4.1 Nodes as competitors

4.2 Simulating Competition