## D. Date Range

Considering the availability of data, including price & volume and other features for later forecasting, we decide to use data from 2007 to 2022. On average, nearly 700 stocks will be added to or deleted from the index. Therefore, the date range gives us nearly 10500 sample. Due to our methodology is to predict the cap change around the rank day. Therefore, we may use any other period data for training or use data from Russell 2000, which have similar rebalance methodology like Russell 3000. The data is relatively sufficient.

In-sample & Out-of-sample:

We use 2007-2020’s data as in-sample for training and use 2021-2022’s data for testing. We may use bootstrap and cross validation in in-sample to improve model performance.

## E. Strategy Detail

Russell 3000 rebalances basing on the rank of stock’s cap on the rank day. Therefore, we are focusing on predicting the change of stock’s cap around the rank day.

**Signal Generation:**

**Portfolio Construction:**

We use signal weighting to constructing portfolio.

**Trade Execution:**

1. Enter the trade

Getting signals one day ahead of the rank day, we will trade the portfolio as early as possible in order to avoid possible liquidity problem after the rank day.

1. Exit the trade

We plan to exit the trade 15 days after the rebalance. The holding period is a parameter requiring further research.

1. Stop Loss

Check with the rank day’s result and rebalance results to adjust portfolio constituents in time.