# Introduction/Abstract:

# Geometric Brownian Motion Derivation:

# Methodology:

# AMD Stock:

## Analysis of Returns:

The distribution of returns on AMD stock is normal with a couple outliers on the right end tail, however the behavior appears to be normal. Similarly, the Q-Q plot shows that the behavior can mostly be described by a normal distribution, but the points that trail off at the end indicate a bit of right skewedness.

Chart, line chart

Description automatically generated

## Multiple One Step Predictions:

The one-step predictions can be applied over a set period to determine how well it forecasts future values of price. In the research paper **[INSERT RESEARCH CITATION]**, it was determined that a training set of 60 days produced the best predictions based on the mean squared value. That experiment is repeated, but with the addition of the mean absolute percentage error as another criterion of forecasting accuracy. 100,000 geometric Brownian motion simulations were generated from training sets of 30 to 100 days each, and the resulting simulations are compared to the actual stock price to obtain the MSE and MAPE.

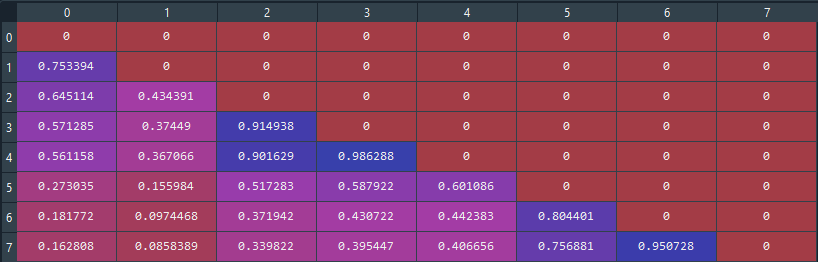
Another interesting methodology that the same research paper conducts is examining the experimental probability of predicting the correct direction of the price change. According to the same research paper, 100-day sets produced the most accurate direction prediction. To perform this experiment, 100,000 one-day simulations will be generated for training sizes from 30 to 100 each. The resulting experimental prices will be subtracted by the true *s(t-1)* price, and the direction will be checked against the true price change direction of *s(t)-s(t-1)*.

The following results are summarized in the figure below:



The expected RMSE and expected MAPE for the 100-day set is the lowest amongst the training sets. This differs from the findings in **[INSERT RESEARCH CITATION**], which concluded a 60-day training set to be the most accurate.

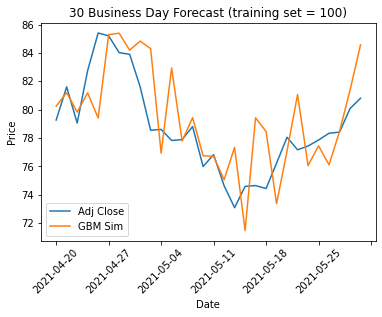
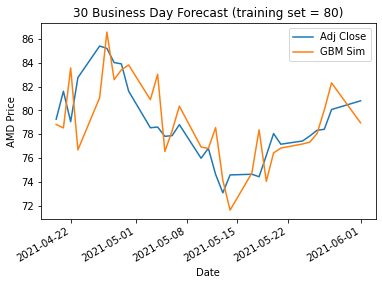
The expected RMSE of each training size are all very similar, so in order to determine if there even is a statistically significant difference between each training size’s expected RMSE, a two-tailed t-test will be performed against each permutation:



The table above displays the calculate p-values of each tested permutation of training sizes. With a critical value of 5%, this test reveals that there’s actually no significant difference between the expected RMSE of the training sizes.

In comparison to the paper’s results of 100 days being the most accurate training size when it comes to predicting the correct direction of the price change, this scenario results in a training size of 80 being the most accurate.

## Sample 30 Business Day Forecasts:

# S&P500 ETF (SPY):

## Analysis of Returns:

The distribution of returns on SPY ETF is a bit skewed to the left. However, it does still hold a general normal shape. The Q-Q plot reinforces this assessment since only a few of the points trail off in the beginning, indicating left skewedness.

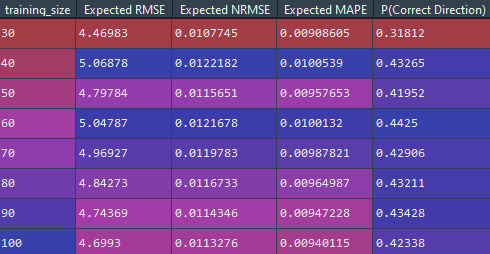
Chart, line chart, scatter chart

Description automatically generated

## Multiple One Step Predictions:

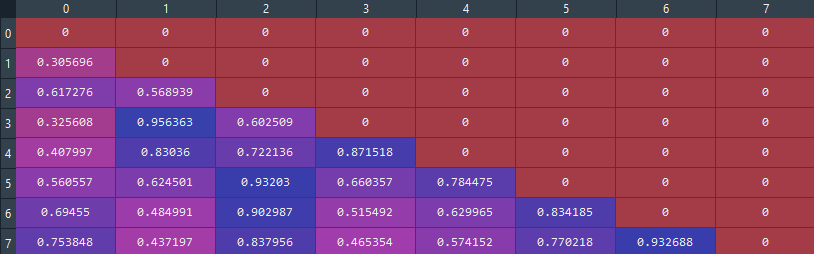
Using varying training set sizes, the size that resulted in the most accurate predictions is a 60-day set with an expected MSE of 11.096. When determined by the expected MAPE, a 30-day set produces the most accurate predictions.

When attempting to determine the best training size to measure direction accuracy, the size that produced the most accurate direction predictions is a 60-day set too.



For the S&P500 ETF, the training size that resulted in the lowest RMSE and MAPE is the 30-day training set. While the 60-day training set resulted in the best probability of correct price movement.

Running the t-test on this set of training data, there is once again no significant difference between the RSME’s of the different training sizes:



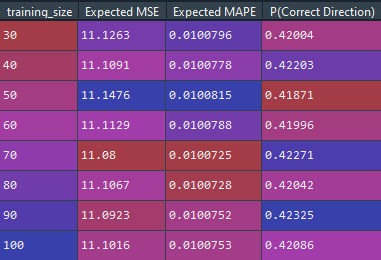
However, when looking at the normalized RMSE (nRSME), it’s approximately three times smaller than AMD’s nRMSE. This seems to be expected given that the ETFs by definition would have less variance than a single stock option since the variance is based on multiple stock options. The smaller variance would result in more predictable and consistent behavior to predict, especially since standard geometric brownian motion assumes a normal distribution.

When it comes to predicting the correct direction of the price change, the ETF results in a training size of 60 being the most accurate.

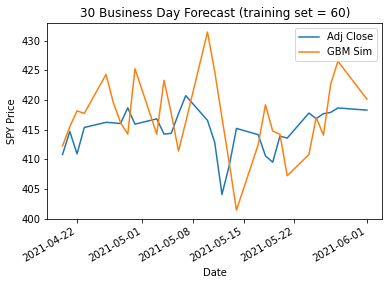
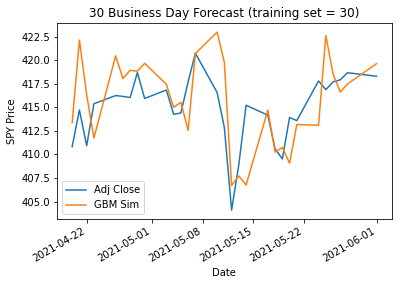
When examining the initial returns distribution, it was noted that the distribution was bit skewed.~~In an attempt to acquire better results, a kernel density estimation distribution was calculated instead.~~

~~Unfortunately, this only resulted in slightly better MSE and probability of predicting a correct directional change. Attempting to diagnose the issue via the distribution shows that the behavior during the test data timeframe has an even greater skew than the KDE of the training data:~~

**KDE was created on the full set of returns – it should only be created on the test data to avoid overfitting**



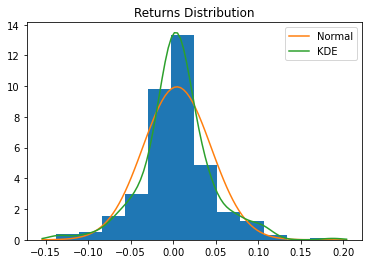
## Sample 30 Business Day Forecasts:



# Bitcoin (BTC):

## Analysis of Returns:

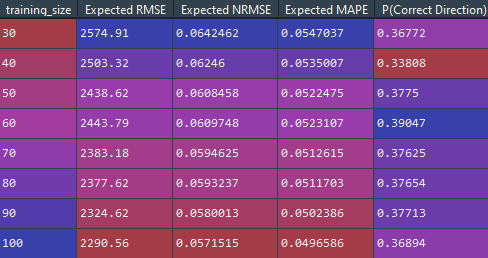
While the general shape and the Q-Q plot can be interpreted as generally normal, the normal distribution does not cover the peaks of the returns enough. Kernel density estimation can be employed to create a distribution that would capture this behavior better.

 Chart, line chart

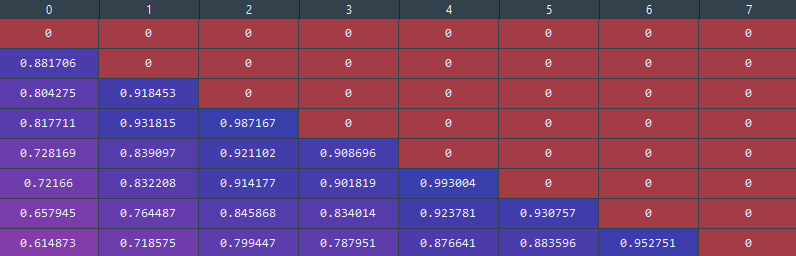
Description automatically generated

## Multiple One Step Predictions:

Training size of 100-days results in the best expected RMSE and MAPE. However, the accuracy of P(Correct Direction) seems low across the board, with the most accurate resulting from a 60-day training set with an experimental probability of 39%. This is possibly due a couple of issues: cryptocurrencies are more volatile than traditional financial assets, and the assumption of normality by the geometric Brownian motion seems to be incorrect based on the returns distribution.



As expected, when running the t-test against the expected RMSE of each permutation of training set size, there is no statistical significance between any of them.



Looking at the expected nRMSE also shows expected results given the volatile nature of cryptocurrencies. The expected nRMSE are higher than that of AMD’s, and six times higher than that of the S&P500’s nRMSE. This means that the forecasting capabilities of standard geometric Brownian motion isn’t as accurate in comparison to its applications to traditional assets where a normal distribution assumption is more acceptable.

~~KDE results in better expected MAPE, and noticeably better P(Correct Direction), but worse expected MSE.~~

**KDE was created on the full set of returns – it should only be created on the test data to avoid overfitting**

A screenshot of a computer

Description automatically generated with medium confidence

## Sample 30 Business Day Forecasts:

