Deciphering Market Dynamics: ARMA-GARCH Modeling for Financial Forecasting

Theo Dawson, Emily Hahn, Zack Horton, Tanner Street
15.072 - Advanced Analytics Edge
December 8, 2023

INTRODUCTION

In the world of volatile financial markets, accurately forecasting the prices of financial instruments poses a significant challenge. Market unpredictability is driven by various factors, including economic indicators, political events, global market trends, and investor psychology. Each of these elements contributes to the continuous change in valuations, making predicting future prices near impossible.

This paper addresses the challenge of financial forecasting by employing advanced statistical analysis and machine learning techniques within R. The primary goal was to fine-tune a time series model that can accurately predict the closing prices of financial instruments up to 14 days in advance. Several hyperparameters and modeling techniques were assessed to better understand the underlying patterns driving market dynamics.

DATA

Data was collected from the publicly available yfinance python package, which offers historical market information from Yahoo Finance. The focal point was on the daily closing prices of the PIMCO Active Bond Exchange-Traded Fund (with ticker BOND). BOND is an ETF encompassing a diverse portfolio of bonds; thus, its trends may indicate broader market trends. The time frame spans five years, from November 2018 through 2023. **Figure 1** in the Appendix displays the full time series.

The data was preprocessed via a log transformation to stabilize variance. Dummy variables for each month and day of the month were implemented to account for potential seasonality, allowing the detection of more underlying trends.

ANALYSIS

Various combinations of regression, autoregressive moving average (ARMA), and generalized autoregressive conditional heteroskedasticity (GARCH) models were trained to generate point predictions and prediction intervals.

Regression Model

The regression models predict the log-transformed closing price based on time (daily basis) and seasonal dummy variables. The time variable is an integer representing the chronological index of the observations. Multiple polynomial degrees on the time variable were tested, ranging from 1 to 4. The dummy variables correspond to each month and day of the month to account for potential seasonality. Each model includes interaction terms between the time variables and dummy variables. **Figure 2** displays the training set valuations along with the corresponding regression predictions.

ARMA Model

ARMA models capture autocorrelation, which is the correlation between a variable's observations over time. These models predict observations based only on past observations (autoregressive component) and past residuals (moving average component). The autoregressive parameter controls how many of the previous observations to use, and the moving average parameter controls how many of the previous residuals to use. These models perform best for time series with a constant mean and no seasonality. There are multiple ways to address mean and seasonality before fitting an ARMA model. In this case, a regression model was applied first; then, the ARMA model was trained on the regression residuals. Therefore, each point prediction is a sum of the regression and ARMA predictions.

The R function 'auto.arima' automatically selects autoregressive and moving average parameters by choosing the pair that minimizes the AIC (in-sample metric akin to adjusted R^2). Using this function for each regression model eliminates the need for manual ARMA parameter testing. **Figure 3** shows the improved regression model with ARMA corrections.

GARCH Model

GARCH models capture heteroskedasticity, which is non-constant variance across a time series. These models predict variance based only on past variances (autoregressive component) and past squared residuals (generalized component). Like with ARMA, each component corresponds to two parameters controlling how many previous values to use. Since GARCH models predict variance, they are used for improving prediction intervals instead of making point predictions. Each prediction interval is the point prediction plus or minus the variance prediction from GARCH multiplied by the 95% z-score.

There are no R functions analogous to 'auto.arima' for GARCH, requiring more manual tuning of parameters. **Figure 4** shows the improved regression model with ARMA corrections and the GARCH upper and lower bounds.

LSTM Model

Deep learning techniques, specifically Long Short-Term Memory (LSTM) models, were briefly explored for modeling time series. Through a series of mathematical gate operations, LSTM learns hundreds of thousands of parameters to identify what pieces of information are important for predicting the future price from current and past data. For this problem, the sequence was chosen to be 20 days, or roughly one month of past information for the model to consider. The LSTM then learns what information from the past is important and what is not. The LSTM only considers data from the past to make predictions and does not consider any of the residual information from the current day per the data architecture defined.

After heuristic hyperparameter tuning, the LSTM was able to achieve a final loss of only 0.00103 on the test set, resulting in a R-Squared value of 0.795 for a ~250 day forecast. This means that LSTM was able to capture approximately 80% of the variance in the data over a 250 day future forecast using ~451,000 parameters. **Figure 8** shows the performance of the LSTM model on the test set.

Evaluation Metrics

Models were evaluated with a multitude of metrics, including Ljung-Box and White tests, mean absolute error (MAE), root mean squared error (RMSE), mean directional accuracy (MDA), and coverage probability (CP).

The Ljung-Box test performs a hypothesis test, where the null hypothesis is that no autocorrelation is present in the time series up to a specified lag. Therefore, if the p-value is below the desired significance level (0.05), the time series contains statistically significant autocorrelation. The ideal model captures all autocorrelation in the time series, so a high p-value is preferred.

Similarly, the White test performs a hypothesis test, where the null hypothesis is that the regression residuals have constant variance. Therefore, if the p-value is below the desired significance level, the residuals are heteroskedastic. The ideal model captures all heteroskedasticity, so a high p-value is preferred.

MAE and RMSE are error measures that calculate the average distances between predicted and observed values. RMSE penalizes large errors more heavily than MAE.

MDA is the probability that the model accurately predicts the direction of change between two sequential observations. An ideal model's MDA is 1, as it perfectly predicts directions of change.

CP is the proportion of actual values contained in the prediction interval. The ideal CP is equal to the confidence of the interval estimate, which in this case, is 95%.

DISCUSSION

Thousands of models were tested, iterating over combinations of polynomial degrees, ARMA orders, and GARCH orders. The models vary significantly in terms of performance. For example, the ratio of the RMSE's standard deviation to mean is greater than one, indicating that there is very high variability in the RMSEs. **Table 1** contains multiple examples from the model evaluation process.

84% of the models capture autocorrelation up to the 14th lag, as indicated by the Box Test. 94% of the models capture heteroskedasticity up to the 14th lag, as indicated by the White Test. The models that fail the Box or White Tests are considered inadequate, as they do not effectively characterize the structure of the time series. Still, most of the models succeeded in accounting for the complex, dynamic structure of a financial instrument.

There is no singular optimal model for predicting the closing prices of financial instruments. Different evaluation criteria should be emphasized for different objectives, as each represents a unique aspect of the model's performance.
For tasks requiring accurate calculations, the best model would have low RMSE and MAE values. If knowledge of market trends is the utmost priority, then models with high MDA scores should be prioritized. Finally, if the objective is to optimize risk mitigation, the optimal models would be those with a CP closest to 95%.

Figure 5 displays the test set performance of a model with strong performance across all metrics. **Figure 6** displays the test set performance of a model with high directional accuracy but high error. **Figure 7** displays the test set performance of a model with low error but low directional accuracy. Each model's exact results are displayed in the first, second, and third rows of **Table 1**, respectively. The figures demonstrate that the models vary significantly in terms of performance, with different models excelling in various metrics.

CONCLUSION

This project tackled the challenge of forecasting prices of financial instruments, with a focus on the PIMCO Active Bond Exchange-Traded Fund. Time series techniques involving regression, ARMA, and GARCH were employed, resulting in thousands of models varying in performance. Selecting the 'best' of these models depends highly on the context, whether the goal is risk mitigation, trend analysis, or price prediction accuracy.

APPENDIX

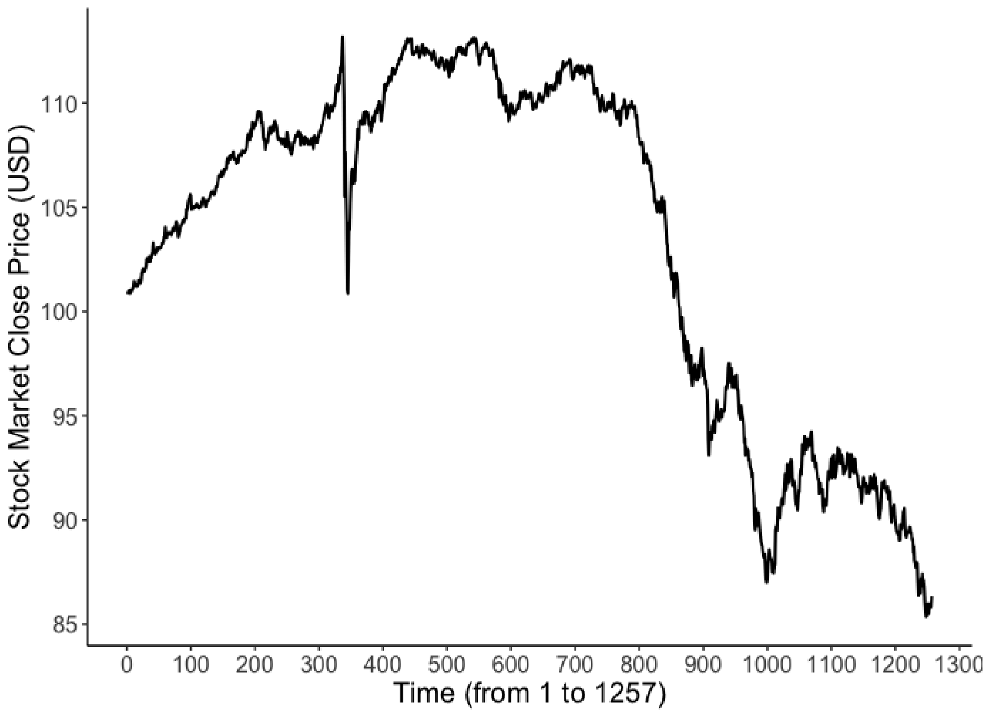

Figure 1: *Valuation of BOND over 5 years*

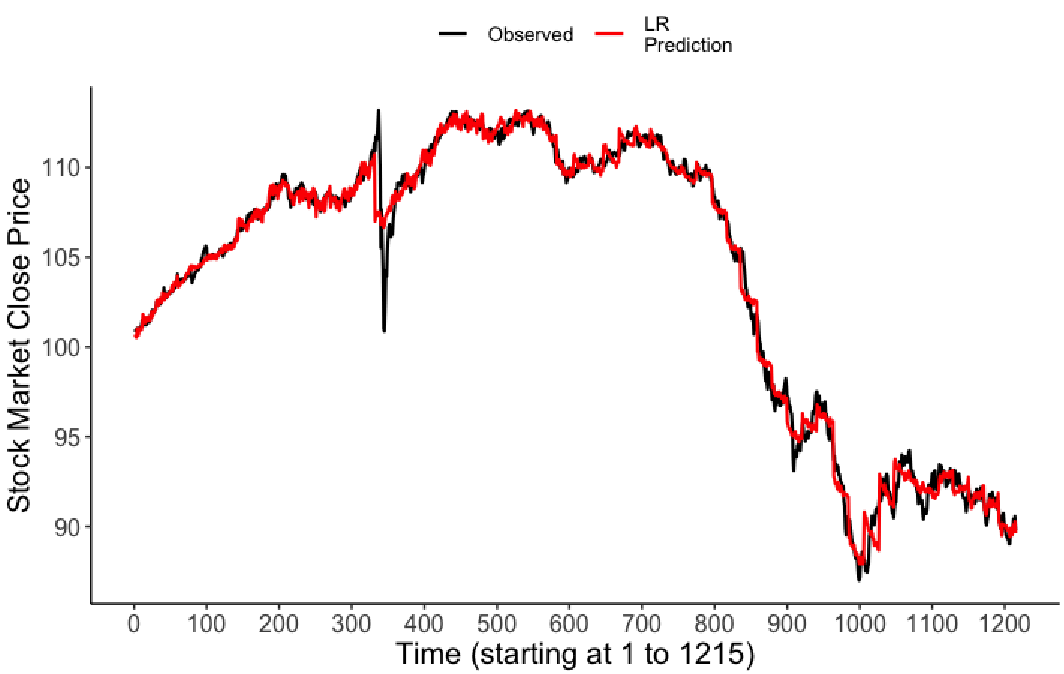

Figure 2: *Linear regression fit to times series*

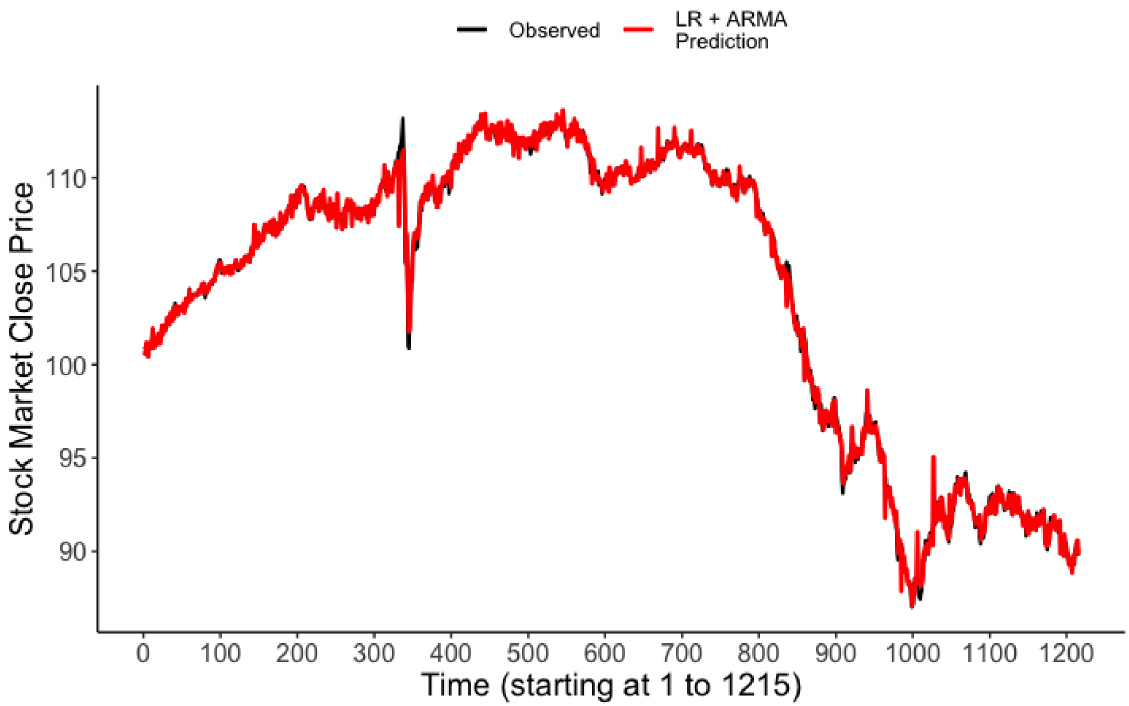

Figure 3*: Linear regression fit to time series with ARMA corrections*

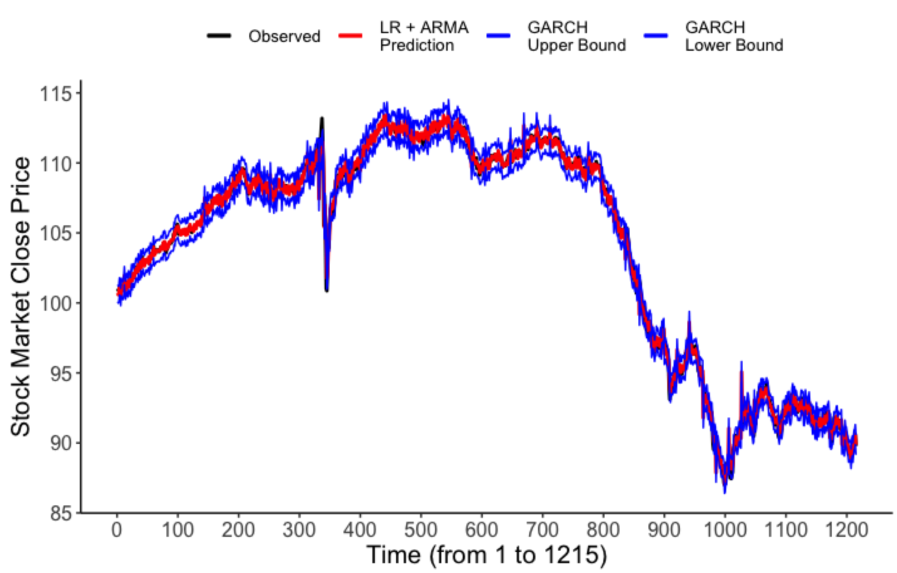

Figure 4*: Linear regression fit to time series with ARMA corrections, with GARCH upper and lower bounds*

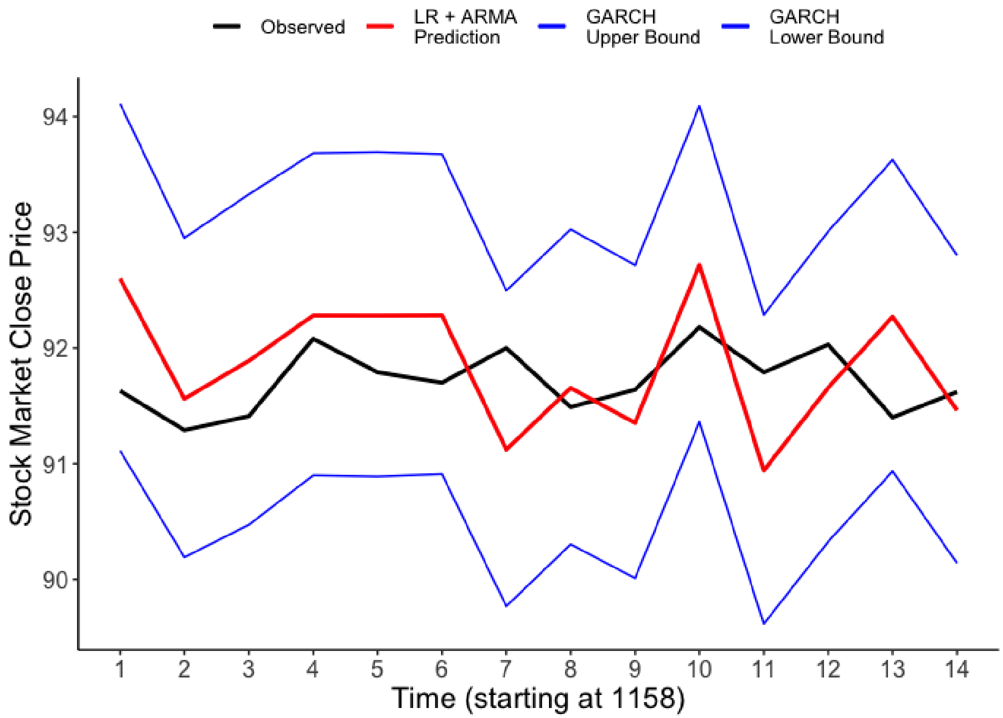

Figure 5: *Model that performs strongly across all metrics; 4th degree polynomial, ARMA(2,1), GARCH(1,1)*

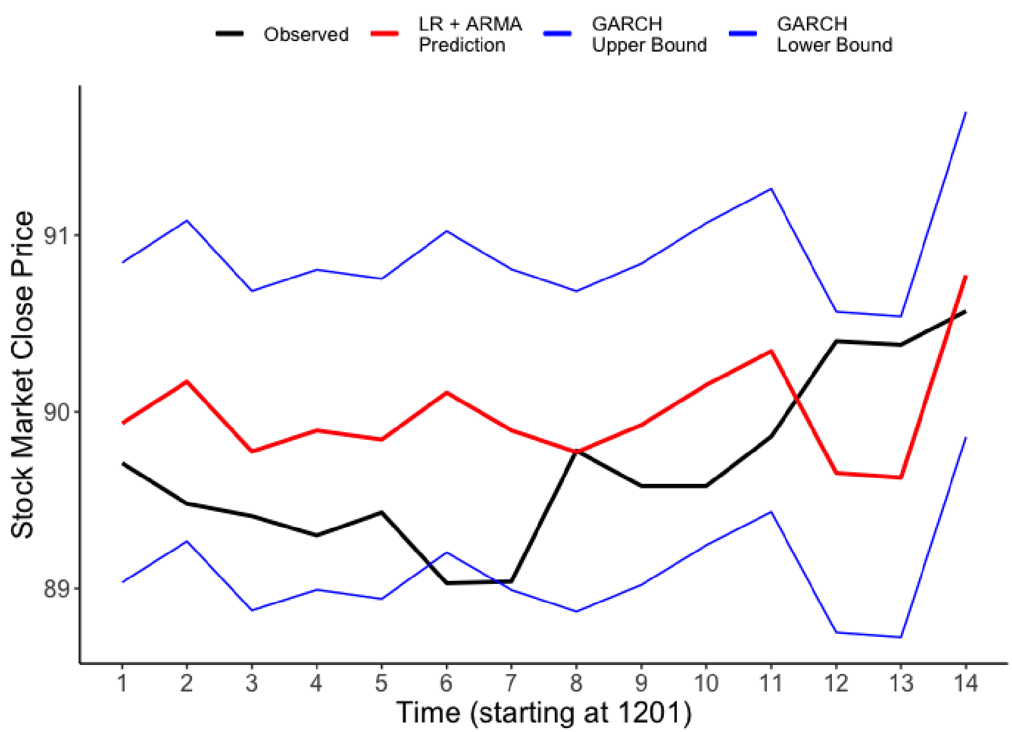

Figure 6: *Model with high directional accuracy but low accuracy; 2nd degree polynomial, ARMA(3,2), GARCH(2,2)*

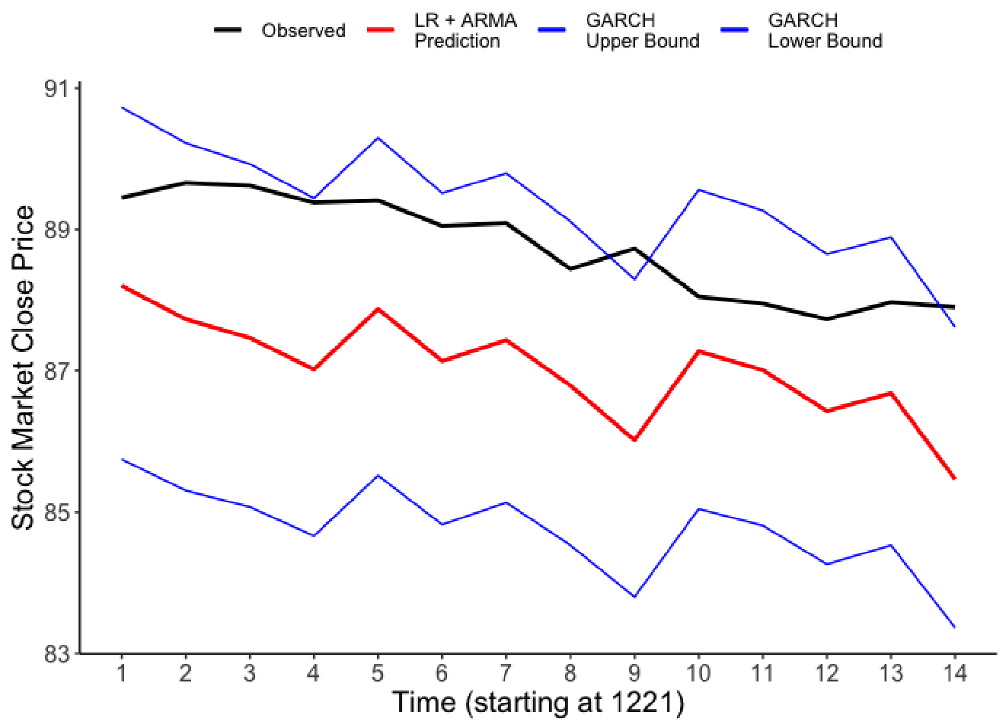

Figure 7: *Model with high accuracy but low mean directional accuracy; 3rd degree polynomial, ARMA(2,1), GARCH(2,2)*

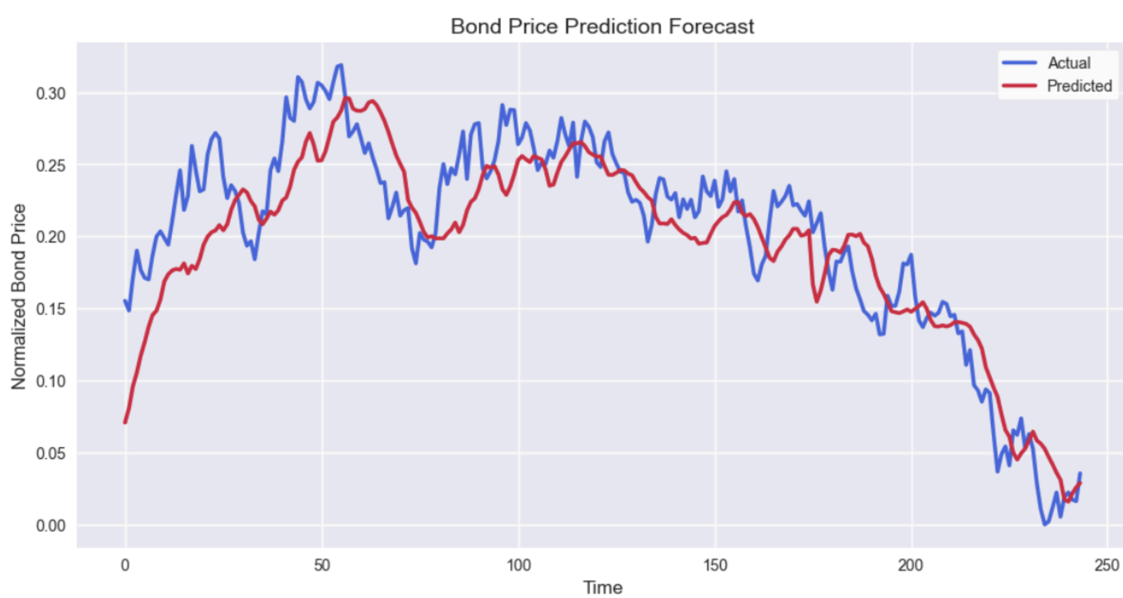

***Figure 8**: LSTM predictions vs. actuals on the test set, ~250 day future forecast*

Table 1: *Examples of model evaluations; ARMA P refers to autoregressive parameter, ARMA Q refers to moving average parameter, GARCH 1 refers to autoregressive parameter, and GARCH 2 refers to generalized parameter; IS refers to in-sample and OS refers to out-of-sample.*

Model Number	Degree Polynomial	ARMA P	ARMA Q	GARCH 1	GARCH 2	Box P-Value	White P-Value	MAE (IS)	RMSE (IS)	MDA (IS)	MAE (OS)	RMSE (OS)	MDA (OS)	Coverage Prob. (OS)
2408	4	2	1	1	1	0.97	1.00	0.0029	0.0044	0.52	0.004	0.005	0.69	0.79
1195	2	3	2	2	2	1.00	1.00	0.0040	0.0061	0.53	0.017	0.018	0.77	0.86
2132	3	2	1	3	1	1.00	1.00	0.0032	0.0050	0.52	0.005	0.006	0.38	0.93
88	1	1	1	1	2	0.81	1.00	0.0052	0.0076	0.51	0.005	0.006	0.54	1.00
1805	3	2	1	1	3	1.00	1.00	0.0034	0.0054	0.50	0.032	0.094	0.77	0.93
1845	3	4	2	2	1	0.88	1.00	0.0034	0.0051	0.53	0.028	0.029	0.31	0.14
2319	3	2	1	3	3	1.00	1.00	0.0032	0.0050	0.52	0.308	0.320	0.62	0.07
1097	2	2	3	2	1	0.98	1.00	0.0034	0.0053	0.53	0.115	0.138	0.62	0.21
1117	2	3	2	2	1	1.00	1.00	0.0039	0.0060	0.54	0.123	0.151	0.69	0.36
635	1	0	2	3	2	0.00	0.99	0.0190	0.0226	0.50	0.046	0.051	0.54	0.43
3042	4	2	1	3	2	0.97	1.00	0.0029	0.0044	0.53	0.008	0.010	0.54	0.57
914	2	2	3	1	2	0.98	1.00	0.0035	0.0053	0.55	0.007	0.010	0.54	0.86
788	2	2	3	1	1	0.99	1.00	0.0036	0.0055	0.55	0.041	0.060	0.62	0.29
2511	4	2	1	1	2	0.97	1.00	0.0029	0.0044	0.52	0.156	0.174	0.85	0.00
2236	3	2	1	3	2	1.00	1.00	0.0035	0.0056	0.52	0.026	0.026	0.62	0.14
475	1	1	1	2	3	0.98	1.00	0.0050	0.0074	0.51	0.008	0.009	0.54	1.00