Forecasts and Present Comparison of ETF Prices

Darian Shane Martos

November 17, 2021

Introduction

From the proposal, I am investigating ETFs and aiming to predict their prices a year into the future, as well as check the predictions against 2021 prices as a baseline. I plan to use ARIMA model fitting for the predictions and parameter estimations, as well as Kalman filters as another method of prediction. If I have time (and am able to catch up with the course lectures), I plan to at least explore LSTM modeling for this data as well.

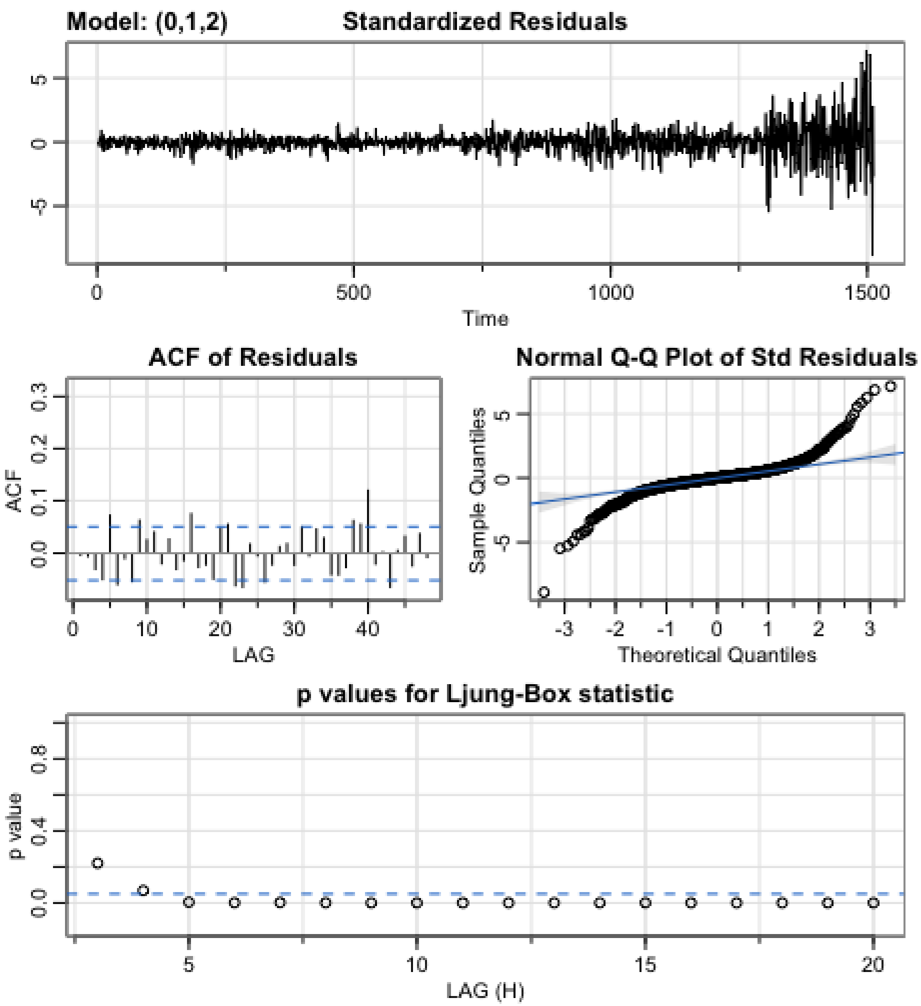
So far, I have explored the ARIMA modeling on two of the ETF datasets (ARKG and QQQ). The ETF data was pulled through a Yahoo finance API, with csvs exported for analysis. The data range from Jan 2015 to Dec 2020. A good majority of the exploration done was over the ARKG dataset, mostly checking over different ARIMA models to fit over. As I'll discuss in the methods section, I'm a bit uncertain in the direction for ARIMA modeling so I will take a pause and go to office hours and talk to staff about ARIMA fitting.

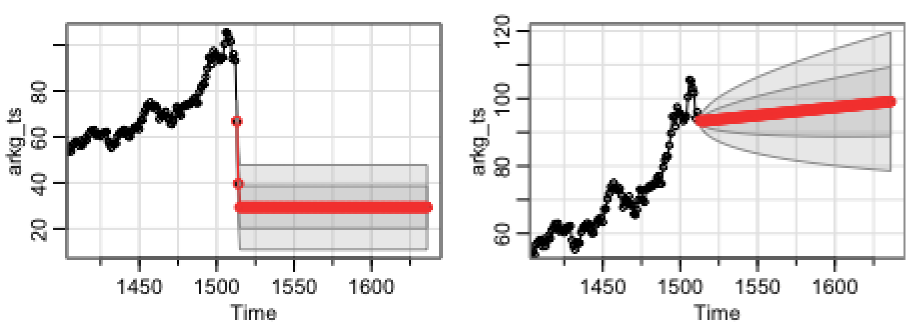
Methods

As seen in experiments, so far I have worked over the ARIMA modeling by fitting ARKG and QQQ data across different order ARIMA models. I follow the textbook process of plotting the data and checking diagnostics to assess better model fit. An obvious challenge currently is extending predictions further than a given h steps (i.e. using previous predictions of new values into future predictions), I have been using the sarima.for function for predictions but find it is flawed in the case of modeling past a certain lag. From the experiments section, this is reflected in the bottom graphic, the left-hand plot shows an MA(0,0,2) model which is the most optimal model with respect to AIC and AIC_c, yet the MA(0,1,2) model functions better works better with sarima.for. My current plans for this week is to resolve any issues with ARIMA modeling, such as improving diagnostics and ensuring proper predictions.

I have done only light exploration of Kalman filtering by reading the reference below, and plan to work more on applying this method (and LSTM modeling or explorations) during Thanksgiving break.

Experiments





As discussed before, I have been following the textbook procedure of model diagnostics and selection in fitting various ARIMA models. For both ARKG and QQQ, I fit five types of ARIMA models of orders $\{(1,1,0),(1,1,1),(0,1,1),(0,1,2),(0,1,4)\}$. A lot of these models were explored passively, I mostly checked the model selection metrics (AIC, etc.) and diagnostics to see what these different order models looked like. From the above images, we have diagnostics for the ARIMA(0,1,2) fit over the ARKG data, and then a sarima.for prediction with an MA(2) model on the left and ARIMA(0,1,2) model on the right. Obviously, a better prediction method in R would be more useful. I may use ARIMA() with predict instead if it appears to work as intended. With the current ARIMA(0,1,2) fit, I had an MSE of $MSE = 155.12$.

Future Directions

As stated above, I intend to work out the issues I currently have with ARIMA modeling. Some current issues to resolve:

- Diagnostics give fairly mixed results. Residual plots appear fairly all over the place, and outliers get heavily skewed on the tails in QQ plots for both datasets. Despite PACF plots indicating both data follow an MA(2) model, the diagnostics for these model fits feel poor overall.
- The Ljung-Box statistic is also significant across all lags for almost all of the models, which is concerning to me.
- A better prediction method than sarima.for is highly desired. From the plot on the bottom left over the predictions given an MA(2) fit, the predicted values are obviously far off.

Any feedback regarding diagnostics would be incredibly useful, I'd like to be able to find the most optimal orders for ARIMA fitting, or also just accept that ARIMA fitting is perhaps not as useful for stock data, especially given the volatility of stock prices within the last two years. After working out these issues, I plan to work in depth on Kalman filtering and LSTM modeling as in above.

Current References Used

1. Stock Price Prediction Using the ARIMA Model
2. Stock Price Prediction with Machine Learning from Project Ideas Page
3. State Space Model and Kalman Filter for Time-Series Prediction