Time Series Analysis of S&P 500 Price Data

Introduction

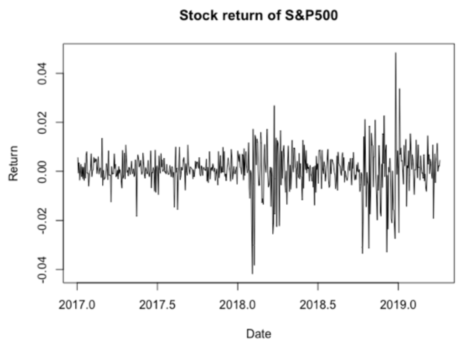
Stock price prediction is always a hot topic as there are valuable applications for different parts of the society. For individuals, it allows people to make more accurate investment in their stock portfolios in order to ensure expected returns are significant and stable. For government sectors, predicting trends in stock price helps frame fiscal and monetary policies in the following years. When observing historical data in the stock market, we can obviously notice that stock prices among adjacent points often exist some sense of correlation that determine how they fluctuate. In this case, time-series analysis becomes useful when we are attempting to interpret events that have happened and will occur on the market.

In this project, we investigated the daily adjusted closing prices of S&P 500 from 2017 to 2019 by employing two types of time-series models: 1) ARIMA(p, d, q) and 2) AR(1) + GARCH(1, 1). Notice that the daily measurements in the stock market exclude weekends, so we only have 5 consecutive trading days in a week. We will examine how the models illustrate the behavior of historical data over the past two years, and how well they predict future trends, fluctuations, and average returns for the next 10 days.

Exploratory Data Analysis and Data Preprocessing

From the initial plot, we can see that there is a generally upward trend at first, followed by a sudden drop at the end of 2018, and finally the increasing trend appeared again. The S&P 500 lost more than 9% in December 2018 due to fears about a tightening monetary policy by the central bank, a slowing economy, and an intensifying trade conflict between the U.S. and China. Also, no missing values are found.

Considering that neither of our models includes trend components, and since we are usually more concerned with change in rate of return, we converted the raw data using a log transformation, followed by one-time differencing. After transformation, the data has no trend and seasonality, and the variability is roughly constant as well, indicating that it has been a stationary time series. Therefore, there is no need for a second differencing, as the more we difference, the more complicated models we attain. The last ten data points in the series were removed during model fitting so that we can evaluate out-of-sample forecasts.



Model Selection

(1) ARIMA(p, d, q)

$$\text{AR(8): } x_t = a_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_8 x_{t-8} + \omega_t,$$
$$\text{where } \omega_t \sim N(0, \sigma_t^2);\ a_0 = \mu(1 - \varphi_1 - \varphi_2 - \cdots - \varphi_8)$$
$$\text{MA(8): } x_t = \mu + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \cdots + \theta_8 \omega_{t-8} + \omega_t, \text{ where } \omega_t \sim N(0, \sigma_t^2)$$

In the previous section, we have made a one-time difference on the time series, so we can determine that d=1. To further derive a suitable ARIMA model for fitting the S&P 500 return data, we first evaluated partial autocorrelation and autocorrelation plots, which are commonly used for determining the values of p and q, respectively. In this case, we can see that as the lag increases, both of them decays to zero asymptotically. However, here we would like to find pure models which are either an autoregressive (AR(p)) or moving average (MA(q)) model, therefore, in one of the plots, we identified the first lag where its subsequent lag begins to converge under the boundary as our cut-off points, p or q, while treated the other plot as tailing off. As a result, by observation, p=8 and q=8 are chosen for the two models, meaning that we would have ARIMA(8, 1, 0) and ARIMA(0, 1, 8) in our modeling work.

(2) AR(1) + GARCH(1, 1)

$$\text{AR(1) -- } x_t = a_0 + \varphi_1 x_{t-1} + u_t$$
$$\text{GARCH(1, 1) -- } \begin{cases} u_t = \sigma_t \varepsilon_t, where\ \varepsilon_t \sim Gaussian \\ \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_t \sigma_{t-1}^2 \end{cases}$$

In the ARIMA model setting, it assumes a constant conditional variance in the time series as a Gaussian white noise with fixed variance is included in the model. Nevertheless, in the stock market, we often notice that when a recent price encounters a volatile state, it is more likely to have a wilder fluctuation in the next following days. To capture such behavior, we can use GARCH(1,1) model to cluster different levels of variance in the data. Alpha (in ARCH term) represents how volatility reacts to new information, while the beta (in GARCH term) represents persistence of the volatility. The sum of alpha and beta is the overall measurement of the rate at which the effect lasts over time.

To exhibit the autoregression property, we mixed the GARCH(1,1) with AR(1) process. There are several ways to derive the best combinations of orders in the AR-GARCH model, but here we choose the one that has been widely used in practice. We first assumed normality in epsilon term, and then relaxed it to a Student's t-distribution to build another AR(1) + GARCH(1) model for comparison.

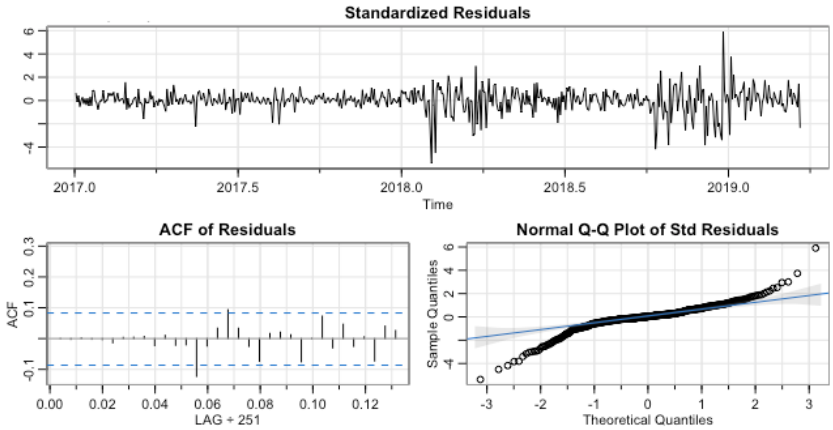
Result

(1) ARIMA Models

From maximum likelihood estimation, the result of model parameterization and diagnostic plots are as follows:

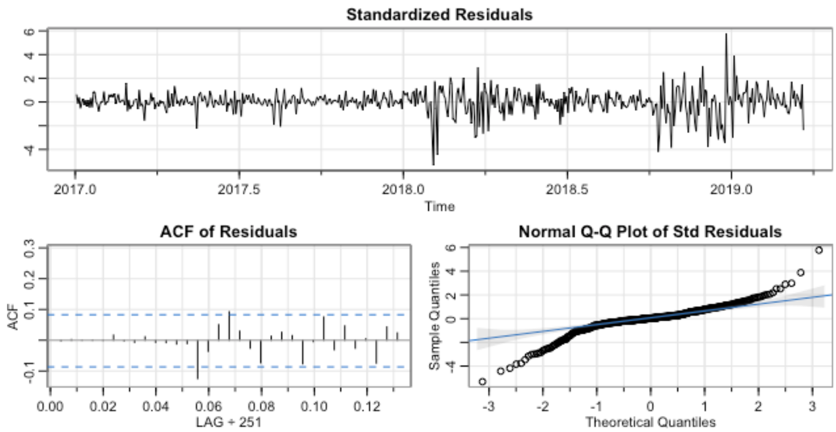
<u>ARIMA(8, 1, 0):</u>

```
Coefficients:
          ar1      ar2     ar3      ar4      ar5     ar6     ar7      ar8   mean
      -0.0235  -0.0394  0.0646  -0.0398  -0.0274  0.0146  0.0432  -0.1209  4e-04
s.e.   0.0423   0.0422  0.0422   0.0423   0.0422  0.0421  0.0421   0.0421  3e-04

sigma^2 = 6.727e-05:  log likelihood = 1889.62
AIC=-3759.23   AICc=-3758.83   BIC=-3716.01
```

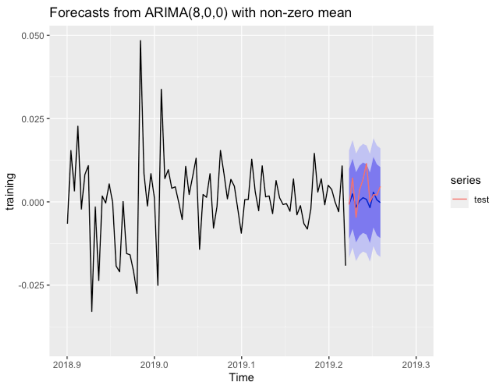


<u>ARIMA(0, 1, 8):</u>

```
Coefficients:
          ma1      ma2     ma3      ma4      ma5      ma6     ma7      ma8   mean
      -0.0197  -0.0478  0.0709  -0.0459  -0.0335  -0.0083  0.0476  -0.1188  4e-04
s.e.   0.0424   0.0425  0.0429   0.0422   0.0433   0.0437  0.0490   0.0411  3e-04

sigma^2 = 6.731e-05:  log likelihood = 1889.43
AIC=-3758.86   AICc=-3758.45   BIC=-3715.63
```



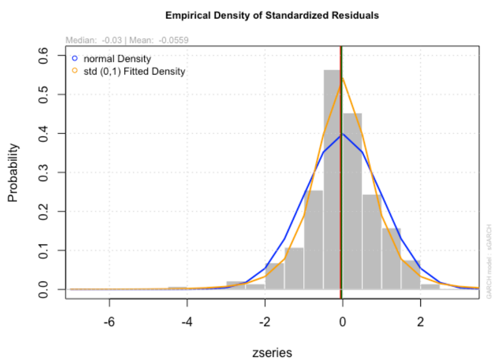
From the coefficient estimation, we found that lag 8 may have the largest influence on the current value. In addition, the positive mean in each model indicates that the average return of the S&P 500 from 2017 to 2019 is positive, and the parameters are significant. Two models fit quite well according to the corresponding diagnostic plots. The standardized residuals plots show that the residuals generally follow a Gaussian white noise, although there exist some outliers greater than 3 standard deviations (reflecting some unusual events in the stock market). The ACF and Q-Q plots indicate that the assumption of normality is quite reasonable, with some exception of outliers described above, causing heavy tails on both sides of Q-Q plots. To determine the final choice in ARIMA, we compared the performance of these two models using information criteria. The AIC, AICc, and BIC favor the pure autoregressive model, so we chose ARIMA(8, 1, 0) for forecasting.



Based on the forecasting plot, we saw that the ARIMA model suggests that the average S&P 500 returns would be positive, slightly fluctuating above zero, and the stock price would keep increasing in the future 10 days. The actual returns (in the orange line) during the time of interest are more volatile and have higher average returns. Overall, the model prediction is similar to the actual returns in the test set, although a more conservative prediction (in the blue line) is provided. The small positive mean derived from the maximum likelihood estimation can justify this conclusion.

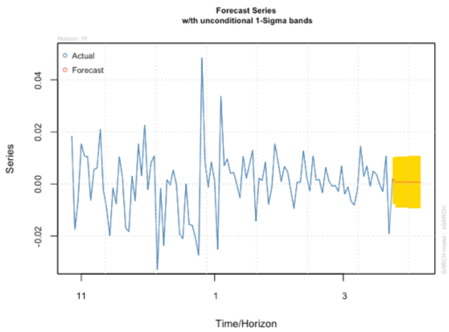
(2) AR-GARCH Models

First, we assumed that the ε_t in the GARCH part follows a normal distribution. Same as what we have seen in the ARIMA model, the mean return is a small positive value. In addition, β_1=0.76 and $\alpha_1 + \beta_1 = 0.96$ suggest that the volatility in the data tends to be persistent. This can be explained by the crisis occurred at the end of 2018. The sum of alpha and beta also indicates that the data is stationary (since it is less than 1). The ACF for standardized residuals looks good as there is no obvious autocorrelation among different points in time. However, the probability distribution shows that the standardized residuals have a higher density than the normal distribution, and the Q-Q plot shows the heavy tails in both extremes. We can improve the model by relaxing the normality assumption in ε_t to a Student's t-distribution.



The parameter estimations in the relaxed version of AR-GARCH model also reflect the persistent volatility in S&P 500 returns from 2017 to 2019. We can see from the diagnostic plots that, compared to the previous assumption of normality, the Q-Q plot and density distribution are much better since the ε_t is closer to the t-distribution with 4 degrees of freedom. All the information criteria favor the AR(1) + GARCH(1,1) with a t-distribution in ε_t, so we will use it to forecast the stock return.

According to the AR-GARCH model, the return on S&P 500 over the next 10 days would be about a constant slightly above zero. Qualitatively, it does not provide an illustration of the fluctuation in the real-world setting, but qualitatively, it successfully predicts the average return in the future 10 days.



Comparison and Discussion

Quantitatively, we found that AR-GARCH (AIC: -7.31, BIC: -7.26) has lower information criteria than ARIMA (AIC: -6.75, BIC: -6.68). However, the forecasting result from ARIMA is more impressive because it successfully reproduces both fluctuations and average future returns, whereas AR-GARCH provides only a horizontal line that only suggests a positive overall return in the future. This can be explained by the unpredictable and complex attributes of the stock market. Although the GARCH model can better capture the nature of variability in stock price, resulting in lower information criteria in the training set, the time we forecasted is right after an unusual, volatile period due to exhaustive factors that are not easy to be considered thoroughly. In some sense, the ARIMA model can better handle such a scenario since it is composed of high orders of autoregressive terms that describe the correlation among points at different times. Perhaps an ARMA-GARCH model with higher orders can make a better prediction for this dataset.