Making Predictions with Google Trend Data

Matt Moocarme

June 16, 2016

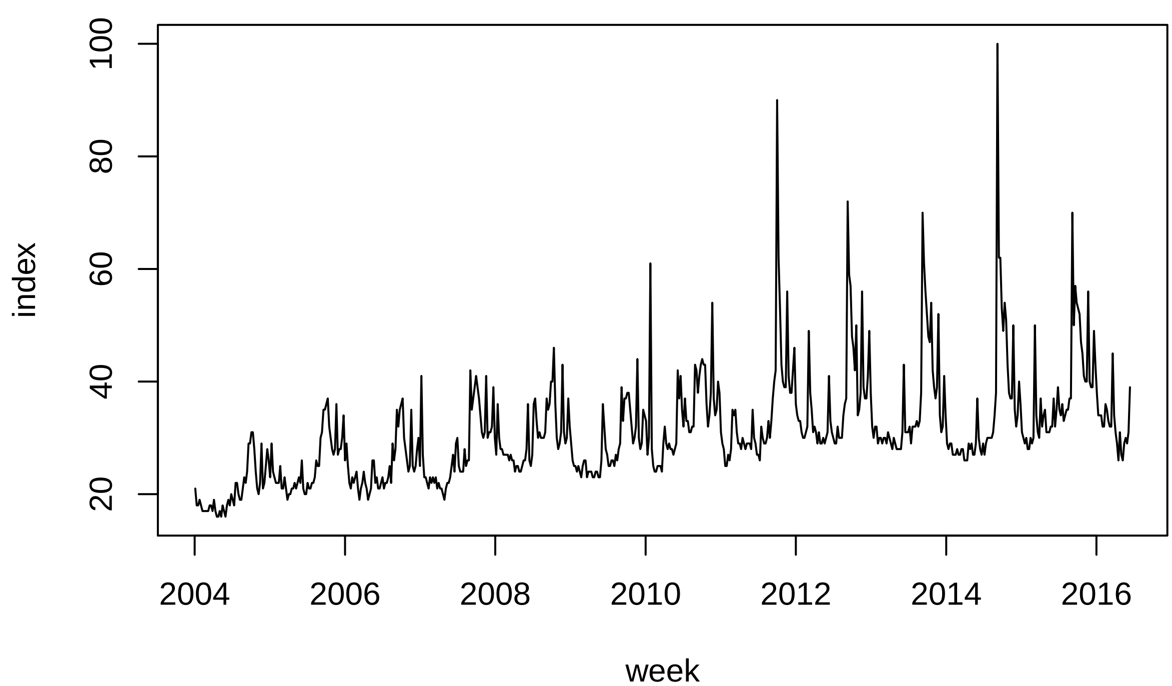
Google Trends as a Predictor

Here I hypothesize that trends in the fluctuation in the stock market can be predicted by the relative changes in amount terms related to the stock index is searched on search engines. For this google trends is useful since it gathers the total number of searches relative to the total search volume in google. For this I used the *googletrend* library to load the data for a particular topic.

I will test this hypothesis on the Apple stock. Apple is very popular and often talked about, if there is no correlation then we cannot reject the null hypothesis: the stock price is independent of relative search volume in google.

In this week I will use the google trend output as a predicter in a time series model. The time series model I will use is an autoregressive intergrated moving average (ARIMA) model, this model will take x number of days of time series data and use it to forecast a given number of days ahead.

The google trend output for keyword 'Apple' in the United States

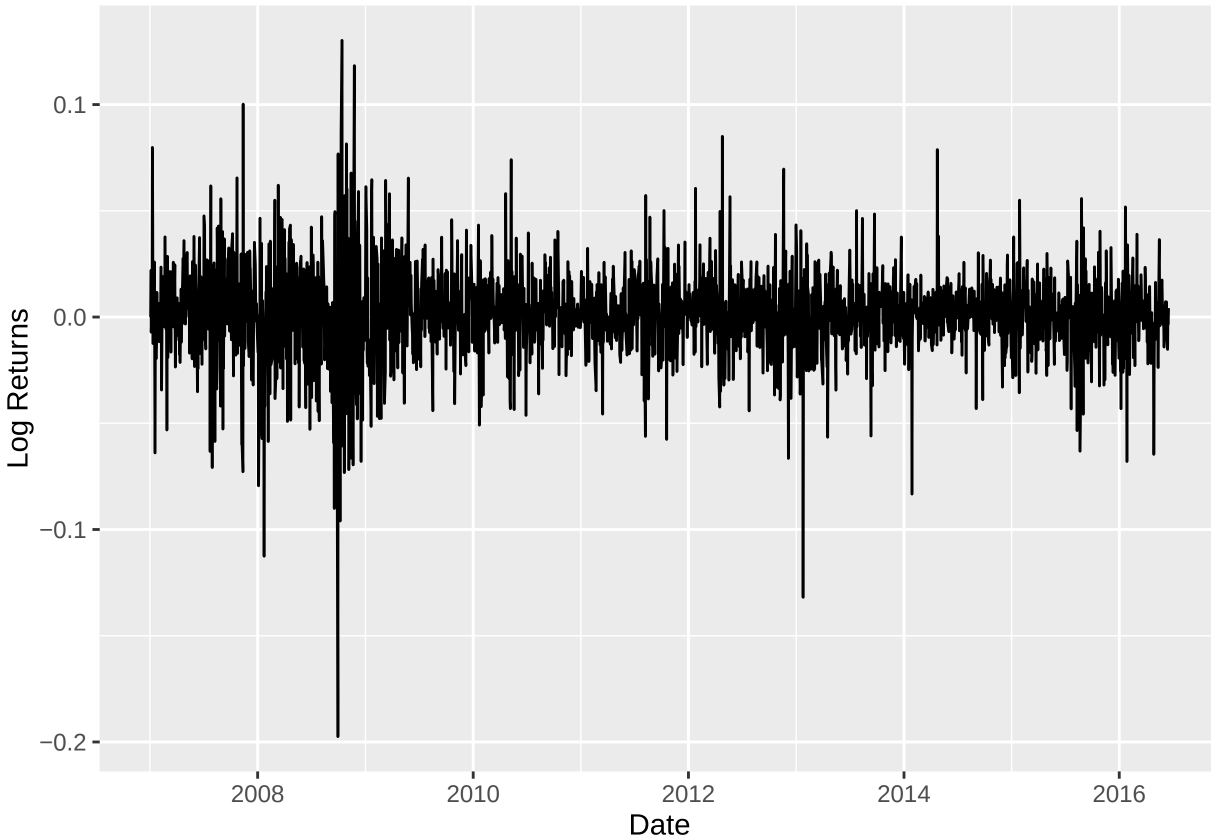


We can see that there is a clear trend in the data. Notably there appears to be some time-dependence, with spikes roughly twice per year.

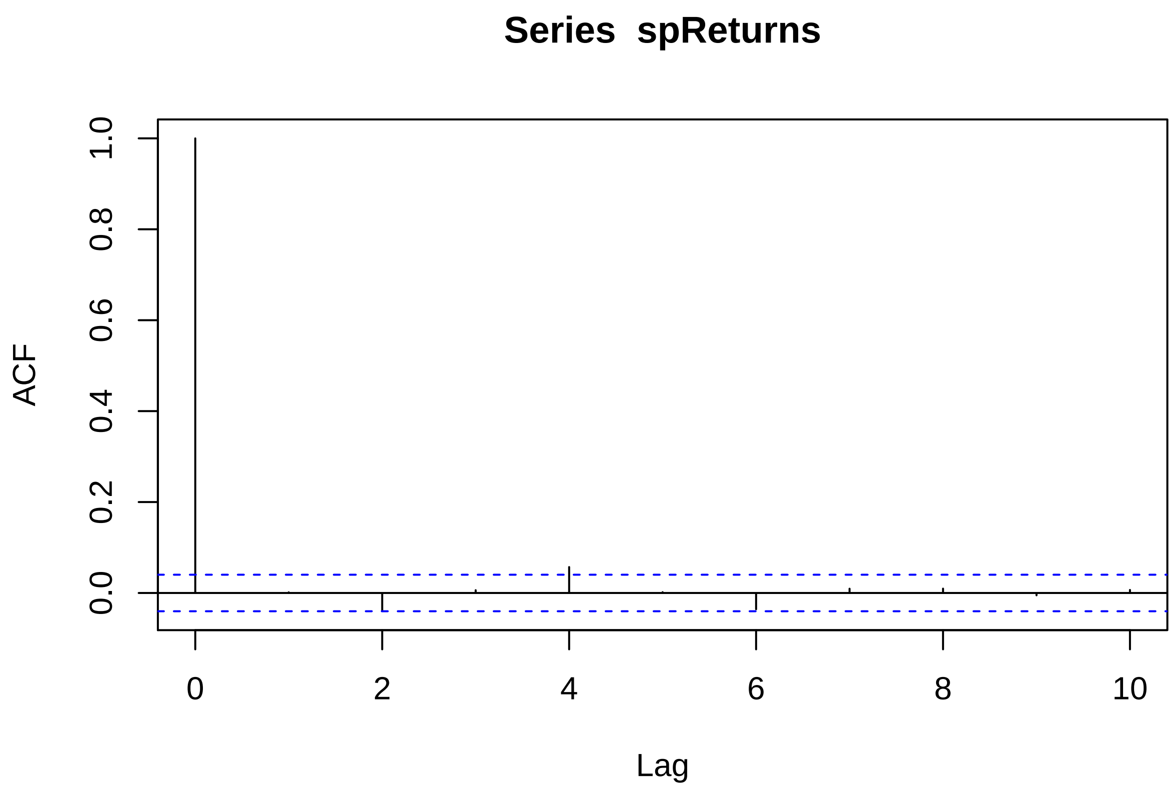
We can also plot the adjusted stock price as a function of time.

```
[1] "AAPL"
```

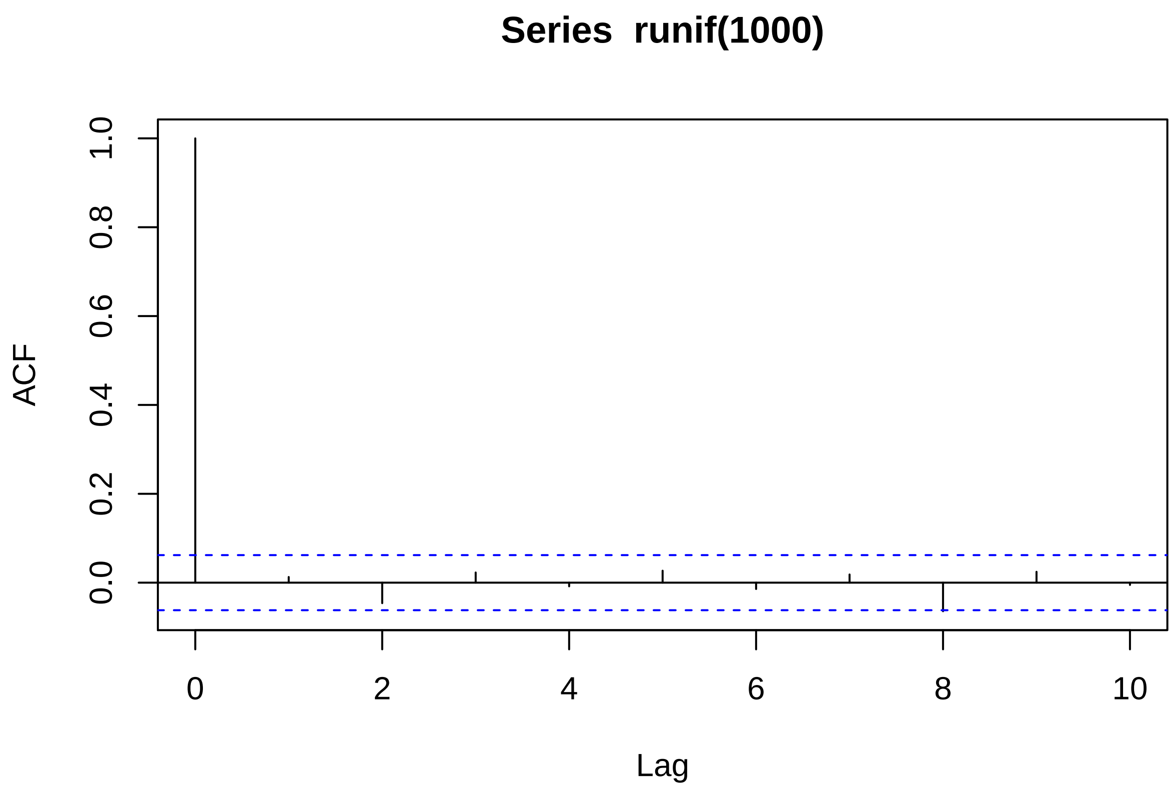
We can also plot the log daily return, given as $r = \log(P_j) - \log(P_{j-1})$, where P_i is the price on a given day, i.



And we can plot the autocorrelation function to see if there is any correlation between the daily rates.



In general there does not seem to much correlation between daily rates, as there are not any spikes above the 0.05 significance line. This indicates that the daily rates are akin to random fluctuations. To prove this we can plot the autocorrelation for uniform random numbers, which should no correlation



We can see that there are higher significant spikes on the random numbers compared to the daily returns. This tells us that it should be quite difficult to forecast the stock market return from looking at previous returns. After 2 days there is a negative spike around 0.05 and a positive spike above 0.05 after 4 days telling us that there is a slightly significant negative correlation after 2 days, a a significant correlation after 4 days in stock market return.

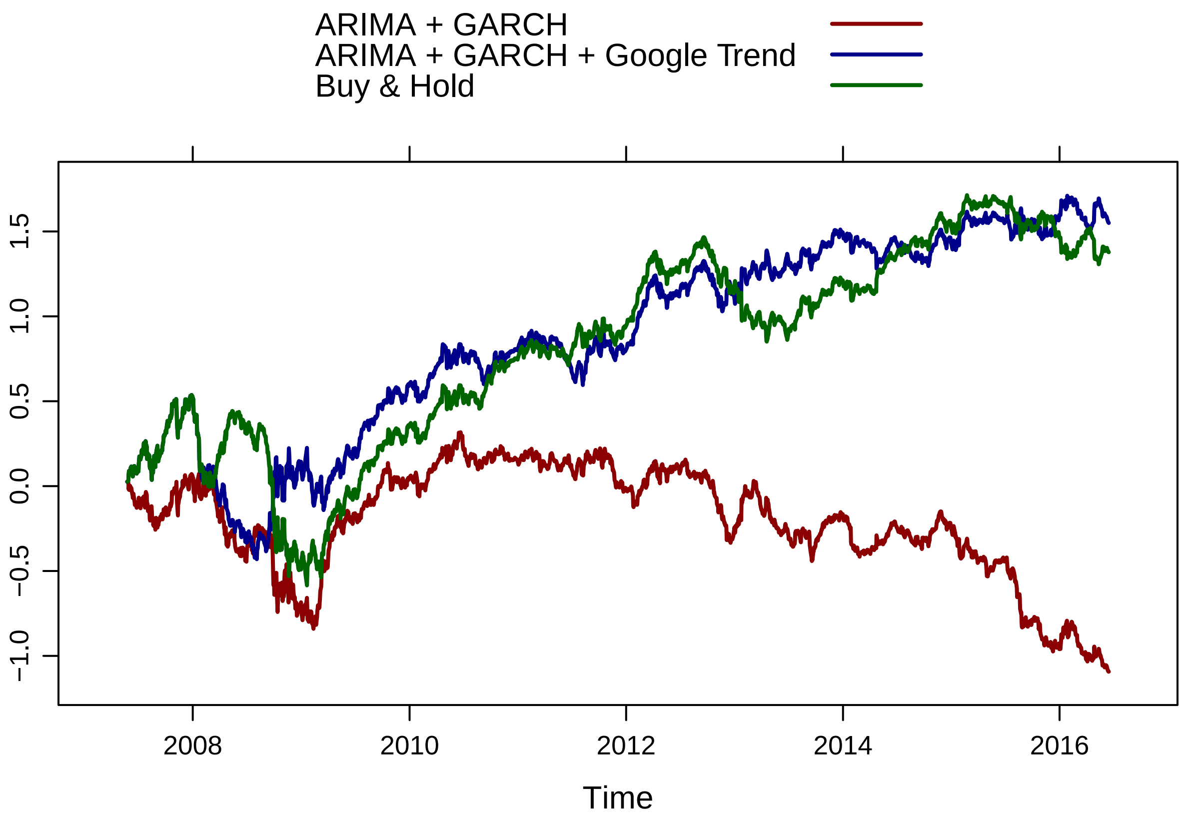
The trend is by no means clear though, which may lend credibility to the idea of looking at google trends as an indicator for stock market returns.

We will attempt to predict the stock market price using the ARIMA model, in which the stock market return is a weighted linear sum of the n last daily stock returns. The ARIMA model has p auro regressive terms, d differencing operations, and q moving average terms. To select the best number of parameters we will run through all combinations in parameter space and pick the best one.

We combine the ARIMA model with the generalized autoregressive conditional heteroscedastic (GARCH) model which models volatility, and looks at how it changes over time. Typically the volatility varies less over time compared to the daily return, so we use larger windows to predict the volatility, such as 100 days.

We run the model, and for positive outcomes, the model predicts the return will be positive, a 1 is outputted, which represents a long position, or buy. Else if the stock return is negative, a -1 is outputted, representing a short postion or sell.

Finally the equity curve is produced, which displays the relative change in value of the asset over time. If the equity curve remained at zero, there would be no change, if the equity curve climbed to 1, the value of the asset doubled.



We can see that the ARIMA-GARCH models do not perform very well, this may be because there is not much autocorrelation in the daily returns of apple stock. When there is serial correlation this model typically performs very well. Moreover including the google trend into the model outperformed the buy and hold strategy, but only slightly.

Moreover, including the google trend data made the model run much faster. Timing the model, without the google trend data the model took 4655 seconds, which is roughly 1 hour 17.6 minutes, whereas including the google trend only took 546 seconds, which is 9.1 minutes. This may be because the threshold for determining whether to buy or sell is acheived faster with the google trend information, compared to without.

Conclusion

We have shown that the including google trend data can influence the performance of a ARIMA-GARCH forecasting model used to predict the daily return on apple stock. This information can not generate larger equity and has a faster run time than the equivalent ARIMA-GARCH model without including google trend data.