

Predicting Stock Trends With Twitter



Sentiment

C. Lee Fanzilli

Advisor – Prof. Dvorak and Prof. Webb

Abstract

With modern advances in technology, the prediction of trends in the stock market and individual stocks has been a hot topic for some time. With the development of various computing methods there have been studies “[demonstrating] that soft computing techniques outperform conventional models in most cases” (Atsalakis et al. 2009). Whenever trying to make predictions in the stock market, we must take into consideration the Efficient Market Hypothesis, or EMH.

Hypothesis

Can we predict future returns of a given stock using Twitter sentiment?

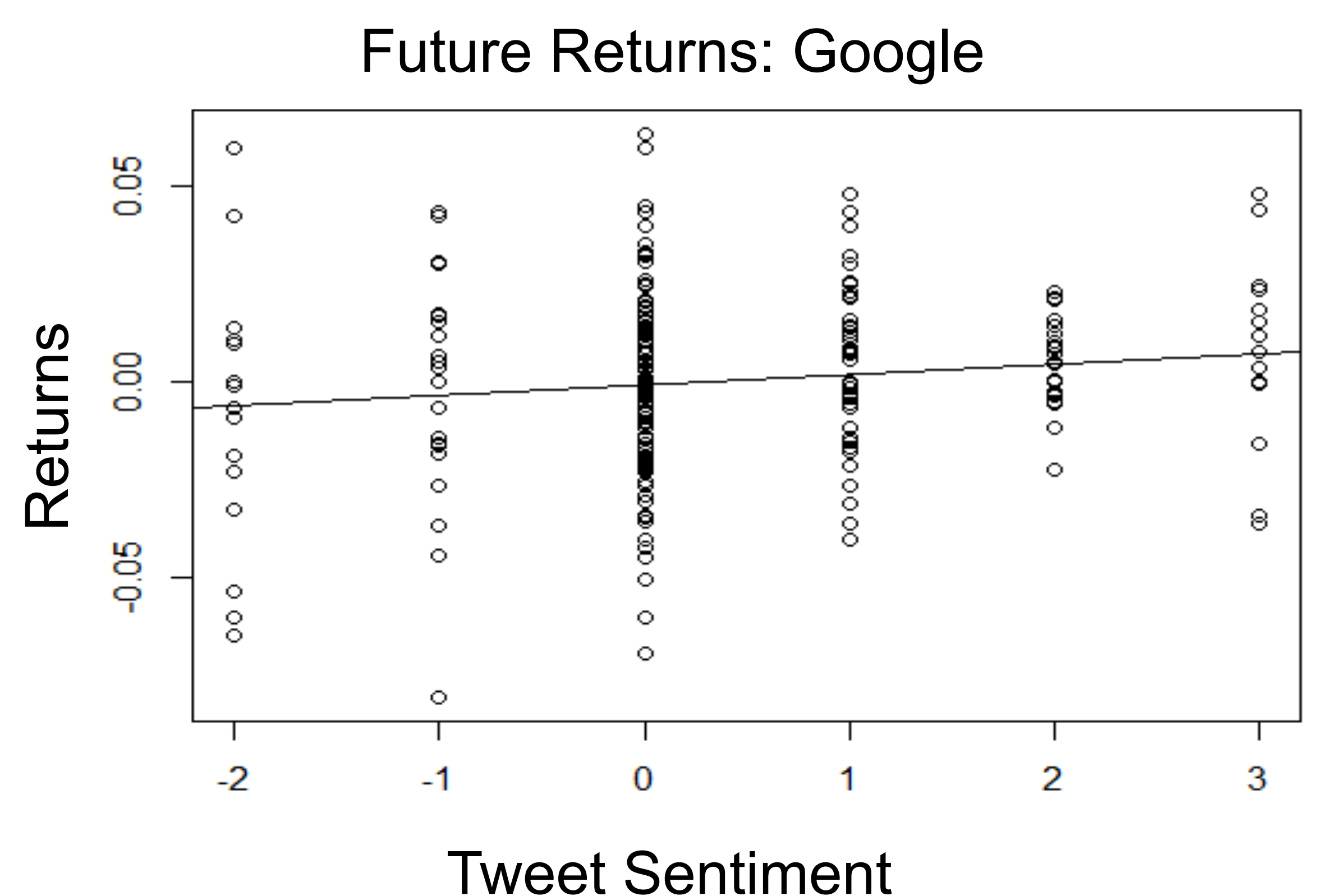
Can we predict daily returns of a given stock using Twitter sentiment?

Do past returns of a given stock predict Twitter sentiment?

Methods

Twitter provides a python API for downloading tweets. We collected a list of 101 tweeters from “Business Insider” and were able to acquire the last 3200 tweets for each individual. In R we combined daily open and closing price data with the twitter sentiment data that was created.

	Return _{t-1}	Return _t	Return _{t+1}
Intercept(AAPL)	t-value = 2.04	t-value = 1.80	t-value = 2.48
Intercept(AMD)	t-value = 0.497	t-value = 0.79	t-value = 0.99
Intercept(GOOG)	t-value = 0.15	t-value = -0.37	t-value = 0.62
Sentiment (AAPL)	t-value = -0.52	t-value = 0.15	t-value = -0.22
Sentiment (AMD)	t-value = 0.12	t-value = -1.023	t-value = -0.51
Sentiment (GOOG)	t-value = -0.09	t-value = 1.28	t-value = 2.61**
R ² (AAPL)	-0.000336	-0.000448	-0.000436
R ² (AMD)	-0.01565	0.000789	-0.01091
R ² (GOOG)	-0.00291	0.001875	0.01745



Results

After running several regressions in R, we concluded that it is unlikely that there is a correlation between Twitter sentiment and future returns of a given stock. Although, we do note that each case tends to vary. With Apple and AMD data, there was no relationship but with Google, we found a weak but significant correlation testing for future returns. This indicates that even though the data points for Google fall far from the regression line, (hence the low R-squared) sentiment provides useful information about future returns for Google.