Financial Econometrics Report

Syed Ahsan Abbas Naqvi, Raghuwansh Raj, Hakon Sandaker

July 2022

1 Scope and Objective

The purpose of this project is to evaluate the impact of Twitter posts by prominent accounts on high frequency stock index and stock price data. In addition, forecasting is made for log-returns and higher moments. Stock data is restricted to FTSE-100 index and top 10 stocks from the FTSE-100. For forecasting on FTSE-100 index, AR(1), GARCH, HAR models were used, while an addition of VAR was made to forecast for the 10 selected tickers.
FTSE-100 is a UK-based index with the majority of tickers inside the UK (80% UK 20% European tickers). Thus, the index might be biased towards micro- and macro economic news for this region of the World. However, there is systematic risk that goes beyond regions, which is what we are trying to capture with the Twitter data, even though Twitter accounts are mainly based in America. It could be events like the Ukraine war or inflation. It will be the sentiment that affects the broader market.

2 Methodology

2.1 Getting Text from Twitter

2.1.1 Identifying Accounts

To get sufficient information and an extensive enough dataset representing the broader population sentiment, we had to query tweets from multiple Twitter groups. Our first group of Tweets was from the Top 50 Most Following Accounts on Twitter (based on Wikipedia). Our second group of tweets was the Top 23 Most Richest Individuals in the World that's on Twitter. Our third group of Tweets was from the US Congress senators.
Since we were quite restricted on time to query data we had to stop scraping data. To further expand the Twitter data down the line we would have tried some Finance News Channels. However, the Finance News Channels can consequently be biased towards political directions and they contain a lot of information as they can be tweeted on a minute basis. It would also be interesting to look at European wealthy individuals on Twitter.

2.1.2 Scraping Tweets

SNScrape, a social networking service scraper in Python, was used to scrap tweets from these individuals. Scraping was done for Tweets published between January 2022 and June, 2022.

2.2 Text Processing

Standard Text pre-processing was conducted on the text of each Tweet such as:

1. Abbreviations Replacement: We replaced abbreviations with their long versions. Contraction replacement e.g. replacing “won’t” with “will not” and “couldn’t” to “could not” etc.

2. Accent Replacement: We replaced accented letters with non-accented versions. Each country uses their own alphabet with different versions of a letter, hence we aimed to standardize the letters. Thus, replacing “á”, “à” and “â” with “a” and so on.

3. Hashtag and ”@” Removal: We removed hashtags from strings but without losing the whole word. We also removed the @ signs wherever people used it to tag other identities on Twitter, since it does not add any extra meaning to the text.

4. Emoji Replacement: We replaced emojis with the descriptions. For example the happy face emoji is replaced by something descriptive e.g., “smiling face, happy”.

5. Unicodes, non-alphabet characters, extra spaces, URL removal

6. Slang Replacement: We replaced slang words with their meaningful replacements. For example, “2 day” is replaced with “today” or “2night” is replaced with “tonight”. We used a predefined slang dictionary file to translate used from a paper [1].

2.3 Language Modelling

Two pre-trained models were used, VADER and GoEmotions Model:
VADER is a lexicon and rule-based feeling analysis instrument that is sensitive to communicated suppositions. It uses a mix of lexical highlights, for example words, which are mostly marked by their semantic directions. The model also gives the degree to which a sentiment is positive or negative. It gives four outputs; Positive Sentiment [-1,1], Neutral Sentiment [-1,1], Negative Sentiment [-1,1], and Compounded Sentiment [-1, 1]. Compounded Sentiment is the average sentiment across both the positive, neutral, and the negative sentiment.
GoEmotions is a recent text sentiment classification made by Google. The method was developed by classifying 55,000 comments from Reddit to 28 categories. It behaves like a BERT (Bidirectional Encoder Representations from Transformers) that applies the bidirectional training of Transformer, a popular

attention model, to language modeling. In essence, the model learns the context of a word based on all of its surroundings: words on its left and right. The output contains 28 classes, describing feelings such as admiration, anger and confusion etc.. The output of the sentiments is on a scale from 0 to 1. We force the model to hand us all classes regardless of their presence in the text. Thus, we might have a lot of sentiments which are not really there.

2.4 Stock Returns, Time Intervals, and Moments

Our FTSE-100 index data are collected ranging from January 2022 to June 2022 and at 1 minute intervals. Rolling variance, which is realized variance in our case, is also ascertained at 5-minute intervals. For tickers, we collected data for top 10 FTSE-100 stocks and at 5-minute intervals. We observed that for most of the stocks, data were not available consistently at all time intervals. For instance, stock X may have missing values at certain 5-minute intervals. After careful observation and analysis, we settled with top 10 stocks in FTSE-100 as they did not have this issue. Direct correspondence from Polygon.io,our data source, revealed that for a ticker, if data was not available for a particular interval, it meant there were no trades executed during that interval.

2.5 Timestamp Adjustments

The Twitter data has timestamps that range across the whole day; while trading may not be taking place when a particular Tweet was published. Thus we aggregated all the Tweets that happened after trading hours i.e., 3pm on each day to the sentiment for the next day since they cannot influence the trades having taken place on the calendar day. Similarly, the Tweets that had been published on the weekends were also aggregated to the next Monday and distributed over whole day. We then performed the inner join of tweets data and the stock market data in such a way that for each timestamp on stock data we collected the Tweets that were published in the last 5-minute interval.

3 FTSE-100 Index Modelling

We start off by taking the log-returns on the index which we also find to be stationary.

3.1 AR(1)

It is reasonable to implement AR(1) model, because log-returns are actually the first difference of log of prices. We impement AR(1) as follows:

$$R_t = \beta_0 + \beta_1 R_{t-1} + \mu_t$$

For generating predictions we conduct single step forecasting.

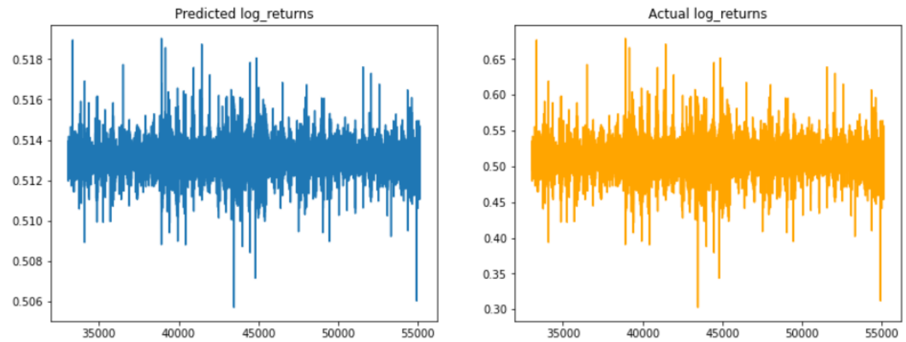

Figure 1: Predicted vs. Actual

As can be seen that forecasts are almost identical to actual returns on the test set which is 40% of the sample, mean squared error (MSE) is 0.00036. However, R-square is very low at 0.5%

3.2 HAR

We then move onto implementing the HAR model to generate forecasts. The idea is that for forecasting, all that matters is the average volatility over the previous week and month. We define the variables as follows:

$$R^w_{t-1} = \frac{1}{5}\sum_{j=1}^{5} R_{t-j}$$

$$R^m_{t-1} = \frac{1}{22}\sum_{j=1}^{22} R_{t-j}$$

$$R_t = \beta_0 + \beta_1 R_{t-1} + \beta_w R^w_{t-1} + \beta_m R^m_{t-1} + \mu_t$$

Again, forecasts are almost identical to actual returns on the test set, mean squared error (MSE) is 0.00034, lower than that of AR(1). However, R-square is again very low at 0.16%.
We then move to add the Twitter sentiment (Vader Compound) component to the model such that:

$$R_t = \beta_0 + \beta_1 R_{t-1} + \beta_w R^w_{t-1} + \beta_m R^m_{t-1} + \beta_T Vader_{t-1} + \mu_t$$

We can see that the predicted values after introducing the tweeter sentiments got more noisy and hence the performance weakened. This explains the negative out of sample R-square value of -0.00697. Corresponding MSE is 0.0029, which is higher than simple HAR and AR(1) implementations.

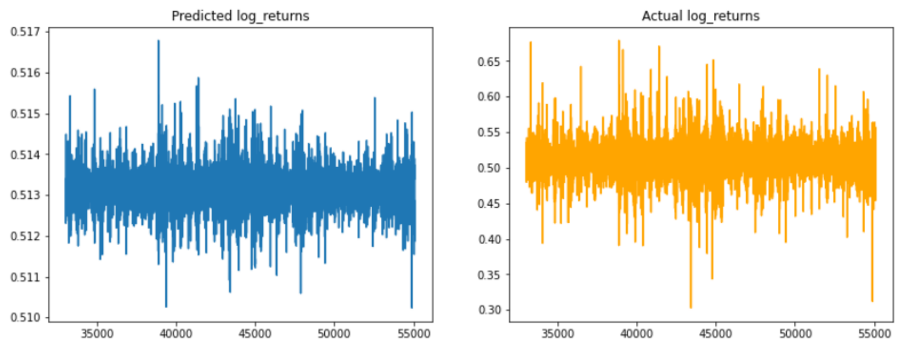

Figure 2: Predicted vs. Actual (HAR)

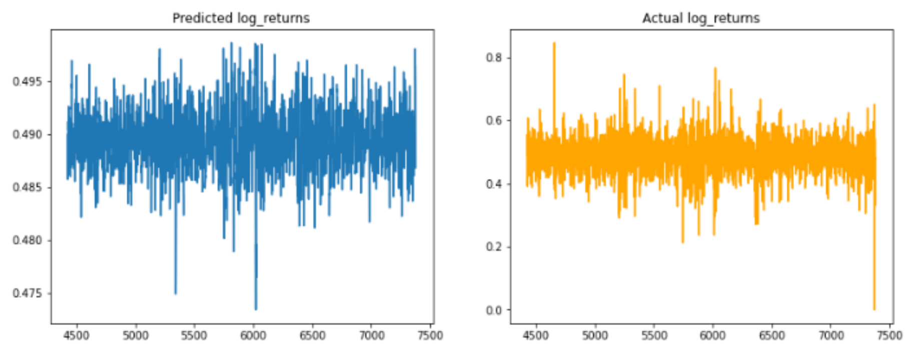

Figure 3: Predicted vs. Actual (HAR with Twitter)

4 Tickers Modelling

Vector Autoregressive model (VAR) was used for the set of 10 stocks. VAR model is a multivariate time series model that relates current observations log-returns with their past observations, past values of other tickers' log-returns and sentiment data. We first implement VAR with other variables as lagged log-returns of other tickers only, and then add sentiment data to make a comparison.

4.1 Granger Causality

The basis behind Vector AutoRegression is that each of the time series in the system influences each other. Using Granger Causality Test, it is possible to test this relationship before even building the model. It tests the null hypothesis that the coefficients of past values in the regression equation is zero. In simpler terms, the past values of time series X do not Granger-cause another series Y. So, if the p-value obtained from the test is less than a chosen significance level, in our case, of 1%, then, we can safely reject the null hypothesis and conclude that X series Granger-causes Y.

	AAL_x	AZN_x	BA_x	BP_x	BTI_x	CUK_x	GSK_x	HL_x	MRO_x	RIO_x
AAL_y	1.000000	0.047700	0.000000	0.001100	0.457000	0.041500	0.036300	0.000000	0.000000	0.000100
AZN_y	0.014700	1.000000	0.027100	0.052000	0.225600	0.042600	0.036800	0.250000	0.236900	0.025000
BA_y	0.000000	0.147800	1.000000	0.023900	0.048400	0.000000	0.304500	0.051400	0.002400	0.114100
BP_y	0.737100	0.303500	0.342800	1.000000	0.134400	0.504400	0.025400	0.000000	0.000000	0.016400
BTI_y	0.065100	0.027000	0.177200	0.000000	1.000000	0.183900	0.005600	0.000400	0.043400	0.006900
CUK_y	0.000100	0.002900	0.000000	0.008100	0.107600	1.000000	0.017100	0.000100	0.000000	0.004100
GSK_y	0.000800	0.000000	0.019100	0.000000	0.000000	0.000700	1.000000	0.221300	0.089500	0.000000
HL_y	0.001500	0.003900	0.039000	0.000000	0.052600	0.001200	0.003400	1.000000	0.000000	0.000000
MRO_y	0.478500	0.007200	0.013000	0.000000	0.004100	0.025200	0.119700	0.032700	1.000000	0.000100
RIO_y	0.824200	0.354800	0.065500	0.002900	0.430700	0.106800	0.443400	0.624100	0.015800	1.000000

Figure 4: p-value Matrix

We can see that 9 stocks Granger-cause at least one stock. Prices were taken as time series, instead of log-returns for this test.

4.2 Cointegration Test

Cointegration test helps to establish the presence of a statistically significant connection between two or more time series, in our case, stock prices, not log-returns. We found out that the order of integration for our series is 1. We find out that none of the stocks are cointegrated as we fail to reject the null hypothesis that there is no cointegration.

4.3 VAR: Ticker Log-Returns Only vs. Log-Returns with Twitter Sentiments

To select the right order of the VAR model, we iteratively fit increasing orders of VAR model and pick the order that produces the lowest AIC.
From the below plot, Figure 2, the AIC drops at lag 2 and then again increase at lag 3. Thus we choose lag 2.
We proceed to check for serial correlation in the series via the Durbin-Watson test. Serial correlation of residuals are used to verify if there is any leftover pattern in the residuals. Existence of correlation in the residuals implies some variation is still left to be explained by the model. We found statistic of 2 for all stocks which meant that there was no significant serial correlation among them. When we forecast using the VAR model, since all series in log-return are stationary, the forecasts are around the mean. In addition, we make forecasts by implementing VAR on all stocks at the chosen lag, and then with the addition of Vader Compound. Both the iterations lead to low RMSE values while with the addition of Twitter sentiment, RMSEs reduce significantly. RMSEs on forecasts of all tickers under both iterations are tabulated below:

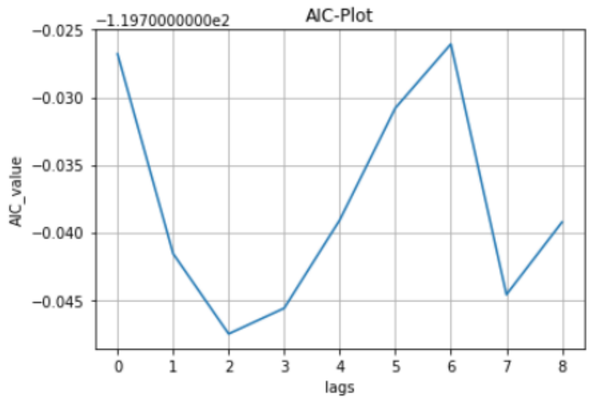

Figure 5: Lags vs. AIC

Ticker	RMSE (Tickers)	RMSE (Tickers + Twitter)
AAL	0.0042	0.0017
AZN	0.0021	0.0001
BA	0.0039	0.0002
BP	0.0029	0.0006
BTI	0.0018	0.0009
CUK	0.0046	0.0011
GSK	0.0015	0.0004
HL	0.0047	0.0034
MRO	0.0046	0.0003
RIO	0.0029	0.0019

4.4 VAR: Ticker Variance Only vs. Variance with Twitter Sentiments

In addition to implementing VAR on log-returns, we implement it on variance as well. For this, we took a range of lags, and found that AIC was minimal at lag 8. However, the moment was stationary without any differencing. In addition, only top 5 stocks were selected for this exercise because of computational expensiveness with 10 stocks.

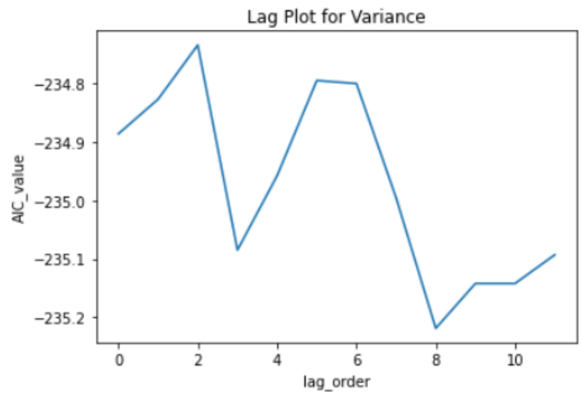

Figure 6: Lags vs. AIC

Forecasts of variances of all stocks are quite close to zero and do not change. This is perhaps because the variances themselves are stationary. In essence, our model is predicting the mean variances. RMSE scores are quite close to zero, and the reason is that both the predicted and actual variances are quite close to zero, as can be observed from the below plots.

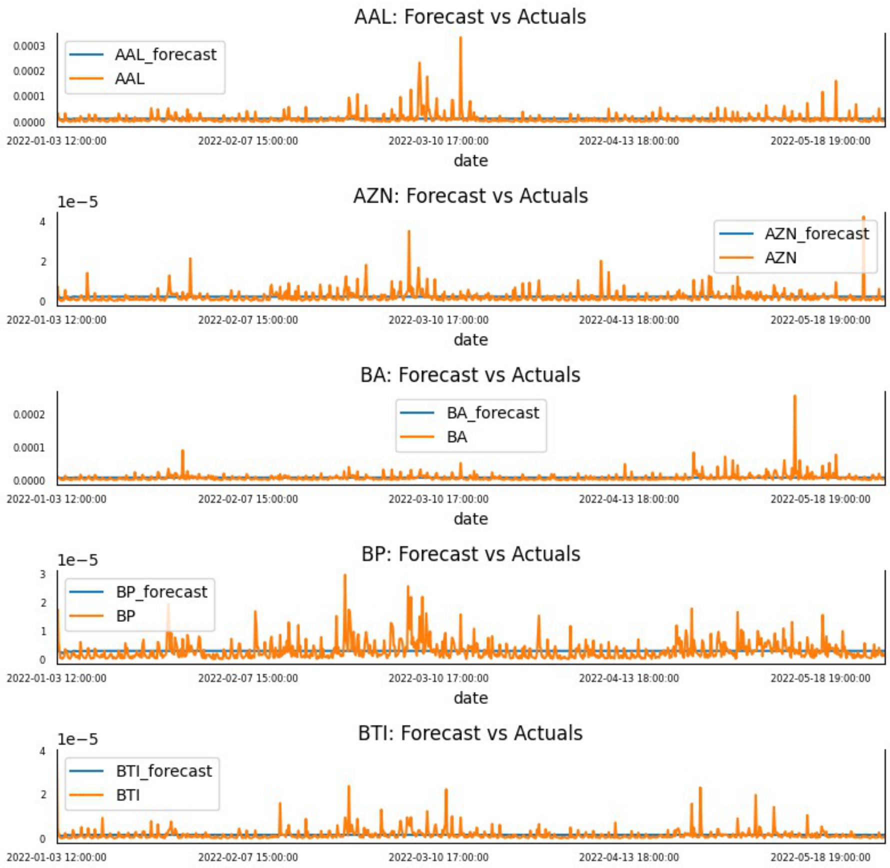

Figure 7: Actual Variance vs. Forecast Variance

A comparison of RMSE scores is presented below:

Ticker	RMSE (Variance)	RMSE (Variance + Twitter)
AAL	0.0	0.0011
AZN	0.0	0.0003
BA	0.0	0.0007
BP	0.0	0.0012
BTI	0.0	0.0027