On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume

Nuno Oliveira[1], Paulo Cortez[1], and Nelson Areal[2]

[1] Centro Algoritmi/Dep. Information Systems, University of Minho, 4800-058 Guimarães, Portugal
nunomroliveira@gmail.com, pcortez@dsi.uminho.pt
[2] Department of Management, School of Economics and Management, University of Minho, 4710-057 Braga, Portugal
nareal@eeg.uminho.pt

Abstract. In this study, we explored data from StockTwits, a microblogging platform exclusively dedicated to the stock market. We produced several indicators and analyzed their value when predicting three market variables: returns, volatility and trading volume. For six major stocks, we measured posting volume and sentiment indicators. We advance on the previous studies on this subject by considering a large time period, using a robust forecasting exercise and performing a statistical test of forecasting ability. In contrast with previous studies, we find no evidence of return predictability using sentiment indicators, and of information content of posting volume for forecasting volatility. However, there is evidence that posting volume can improve the forecasts of trading volume, which is useful for measuring stock liquidity (e.g. assets easily sold).

Keywords: Microblogging Data, Returns, Trading Volume, Volatility, Regression.

1 Introduction

Mining microblogging data to forecast stock market behavior is a very recent research topic that appears to present promising results [3,14,7,9]. In such literature, it is argued that a model that accounts for investor sentiment and attention can potentially be used to predict key stock market variables, such as returns, volatility and volume. Several arguments support this approach. For example, some studies have shown that individuals' financial decisions are significantly affected by their emotions and mood [10,8]. Also, the community of users that utilizes these microblogging services to share information about stock market issues has grown and is potentially more representative of all investors. Moreover, microblogging data is readily available at low cost permitting a faster and less expensive creation of indicators, compared to traditional sources (e.g. large-scale surveys), and can also contain new information that is not present in historical quantitative financial data. Furthermore, the small size of the message (maximum 140 characters) and the usage of cashtags (a hashtag identifier for financial

L. Correia, L.P. Reis, and J. Cascalho (Eds.): EPIA 2013, LNAI 8154, pp. 355–365, 2013.

stocks) can make it a less noisy source of data. Finally, users post very frequently, reacting to events in real-time and allowing a real-time assessment that can be exploited during the trading day.

Regarding the state of the art, in 2004 the landmark paper of Antweiler and Frank [2] studied more than 1.5 million messages posted on Yahoo! Finance, suggesting the influence of post messages in the modeling of financial stock variables. More recently, Bollen et al. [3] measured collective mood states (e.g. "positive", "negative", "calm") through sentiment analysis applied to large scale Twitter data, although tweets were related with generic sentiment (e.g. "Im feeling") and not directly related to stock market. Still, they found an accuracy of 86.7% in the prediction of the Dow Jones Industrial Average daily directions. Sprenger and Welpe [14] have used sentiment analysis on stock related tweets collected during a 6-month period. To reduce noise, they selected Twitter messages containing cashtags of S&P 100 companies. Each message was classified by a Naïve Bayes method trained with a set of 2,500 tweets. Results showed that sentiment indicators are associated with abnormal returns and message volume is correlated with trading volume. Mao et al. [7] surveyed a variety of web data sources (Twitter, news headlines and Google search queries) and tested two sentiment analysis methods to predict stock market behavior. They used a random sample of all public tweets and defined a tweet as bullish or bearish only if it contained the terms "bullish" or "bearish". They showed that their Twitter sentiment indicator and the frequency of occurrence of financial terms on Twitter are statistically significant predictors of daily market returns. Oh and Sheng [9] resorted to a microblogging service exclusively dedicated to stock market. They collected 72,221 micro blog postings from `stocktwits.com`, over a period of three months. The sentiment of the messages was classified by a bag of words approach [12] that applies a machine learning algorithm J48 classifier to produce a learning model. They verified that the extracted sentiment appears to have strong predictive value for future market directions.

While this literature favors the use of microblogging data to forecast stock market behavior, the obtained results need to be interpreted with caution. According to Timmermann [16], there is in general scarce evidence of return predictability. And most of these studies do not perform a robust evaluation. For instance, only modeling (and not prediction) was addressed in [2] and [14], while very short test periods were performed in [3] (19 predictions), [7] (20 and 30 predictions) and [9] (8 predictions). Also, several works, such as [14][9], require a manual classification of tweets, which is prone to subjectivity and is difficult to replicate.

The main goal of this paper is to overcome the limitations of previous studies, by adopting a more robust evaluation of the usefulness of microblogging data for predicting stock market variables. Similarly to [9], and in contrast with other studies [2][3][14][7], we use StockTwits data. Such resource is more interesting, when compared with Twitter, since it is a social service specifically targeted for investors and traders. Also of note, we analyze a much larger data period (605 days) and adopt a robust fixed-sized rolling window (with different window sizes),

leading to a test period that ranges from 305 to 505 predictions. Moreover, rather than predicting direction, such as performed in [9] (which is of lower informative value), we adopt a full regression approach and predict three market variable values for five large US companies and one relevant US index. Aiming to replicate the majority of previous works [14][7][3] we adopt automated methods, using the multiple regression as the base learner model and test five distinct predictive models (including a baseline that does not use microblog data). However, in contrast with previous studies (e.g. [3] and [7] only use MAPE), we use two error metrics (MAPE and RMSE) and adopt the equality of prediction statistical test to make inferences about the statistical significance of the results [4].

The rest of the paper is organized as follows. Section 2 describes the data and methods. Next, Section 3 presents and discusses the research results. Finally, Section 4 concludes with a summary and discussion of the main results.

2 Materials and Methods

2.1 StockTwits and Stock Market Data

Data was collected for five large US companies and one index: Apple (AAPL), Amazon (AMZN), Goldman Sachs (GS), Google (GOOG), IBM (IBM) and Standard and Poor's 500 Index (SPX). These stocks were chosen because they have a substantial posting volume on StockTwits and Chicago Board Options Exchange (CBOE) provides their implied volatility indexes. Therefore, we can process a significant amount of microblogging data that can be more indicative of investors' level of attention and sentiment on these stocks, and use a good volatility estimate. For each stock, we retrieved StockTwits and stock market data from June 1, 2010 to October 31, 2012, in a total of 605 trading days.

StockTwits (`stocktwits.com`) is a financial platform with more than 200,000 users that share information about the market and individual stocks. Similarly to Twitter, messages are limited to 140 characters and consist of ideas, links, charts and other data. We selected StockTwits content because it is exclusively about investing, resulting in a less noisy data set than collecting from a more generalist microblogging service. Messages were filtered by the company $TICKER tag, i.e., $AAPL, $AMZN, $GOOG, $GS, $IBM, $SPX. A $TICKER tag (cashtag) is composed by the company ticker preceded by the "$" symbol. We collected a larger number of messages (Figure 1), ranging from a total of 7283 (IBM) to 364457 (AAPL) tweets.

The stock market variables here considered are daily returns, volatility and trading volume. Price and volume data were collected from Thompson Reuters Datastream (`http://online.thomsonreuters.com/datastream/`) and volatility, as measured by their indexes, was collected from the Chicago Board Options Exchange CBOE (`http://www.cboe.com/micro/EquityVIX/`).

Market returns measure changes in the asset value. We used the adjusted close prices to calculate returns. Adjusted close price is the official closing price adjusted for capital actions and dividends. We computed market returns (R_t) using the following formula [7]:

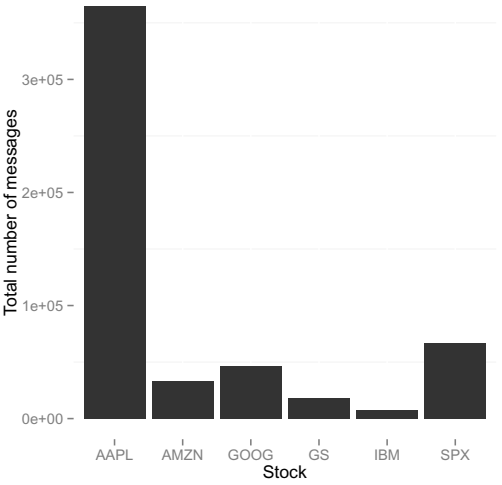

Fig. 1. Total number of StockTwits collected messages per stock

$$R_t = \ln(P_t) - \ln(P_{t-1}) \tag{1}$$

where P_t is the adjusted close price of day t and P_{t-1} is the adjusted close price of the preceding day. Returns provide useful information about the probability distribution of asset prices. This is essential for investors and portfolio managers as they use this information to value assets and manage their risk exposure.

Volatility (σ_t, for day t) is a latent measure of total risk associated with a given investment. Volatility can be estimated using different approaches. Previous studies have found that the model-free implied volatility index in an appropriate estimator of volatility [6]. Estimates of volatility are essential for portfolio selection, financial assets valuation and risk management.

Trading volume (v_t) is the number of shares traded in each day during a trading session. Volume can be used to measure stock liquidity, which in turn has been shown to be useful in asset pricing as several theoretical and empirical studies have identified a liquidity premium. Liquidity can help to explain the cross-section of expected returns [1].

2.2 Regression Models

Similarly to previous works [14][7][3], we adopt a multiple regression model, which is less prone to overfit the data:

$$\hat{y} = f(x_1, ..., x_I) = \beta_0 + \sum_{i=1}^{I} \beta_i x_i \tag{2}$$

where $\hat{y}$ is the predicted value for the dependent variable y (target output), x_i are the independent variables (total of I inputs) and $\beta_0, \ldots, \beta_i$ are the set of parameters to be adjusted, usually by applying a least squares algorithm. Due to its additive nature, this model is easy to interpret and has been widely used in Finance. Moreover, the learning process with such model is very fast, allowing an extensive experimentation of different input variables and a more robust evaluation, with different combinations of training set sizes (Section 2.3).

In this work, we test models that are quite similar to most previous works [14][7][3][9]. This means that for predicting returns we use sentiment data (i.e. the investors opinion about a given stock), while for predicting volatility and volume we explore posting volume indicators (which is a measure of attention). For each financial target, we test five different regression models, aiming to mimic the models proposed in previous works and also testing new variations.

As sentiment indicators, for each stock we count the daily number of messages that contain (case insensitive) the words "bullish" ($Bull_t$, for day t) or "bearish" ($Bear_t$) [7]. Using these two indicators, we compute other variables: bullishness index ($bind_t$) [2][14][9], Twitter Investor Sentiment (TIS_t) [7] and a ratio of TIS ($RTIS_t$, proposed here). These are given by:

$$\begin{array}{ll} bind_t = & \ln\left(\frac{1+Bull_t}{1+Bear_t}\right) \\ TIS_t = & \frac{Bull_t+1}{Bull_t+Bear_t+1} \\ RTIS_t = & \frac{TIS_t}{TIS_{t-1}} \end{array} \tag{3}$$

The five tested regression models for predicting the returns are:

$$\begin{array}{ll} \hat{R}_t = f(R_{t-1}) & (\textbf{M1}, \text{ baseline}) \\ \hat{R}_t = f(R_{t-1}, \ln(TIS_{t-1})) & (\textbf{M2}) \\ \hat{R}_t = f(R_{t-1}, \ln(RTIS_{t-1})) & (\textbf{M3}) \\ \hat{R}_t = f(bind_{t-1}) & (\textbf{M4}) \\ \hat{R}_t = f(\ln(TIS_{t-1})) & (\textbf{M5}) \end{array} \tag{4}$$

Regarding the posting volume, we measure two indicators: n_t, the daily number of tweets (for day t); and $MA_t = \frac{1}{5}\sum_{k=t-5}^{t} n_k$, the moving average (when considering the last five days). Similarly to predicting the returns, and given that in some cases very high values are found for the target, we opt for modeling the natural logarithm values for volatility and trading volume. For predicting volatility, the tested models are:

$$\begin{array}{ll} \ln(\hat{\sigma}_t) = f(\ln(\sigma_{t-1})) & (\textbf{M1}, \text{ baseline}) \\ \ln(\hat{\sigma}_t) = f(\ln(\sigma_{t-1}), \ln(n_{t-1})) & (\textbf{M2}) \\ \ln(\hat{\sigma}_t) = f(\ln(\sigma_{t-1}), \ln(n_{t-1}), \ln(n_{t-2})) & (\textbf{M3}) \\ \ln(\hat{\sigma}_t) = f(\ln(\sigma_{t-1}), \ln(MA_{t-1})) & (\textbf{M4}) \\ \ln(\hat{\sigma}_t) = f(\ln(\sigma_{t-1}), \ln(MA_{t-1}), \ln(MA_{t-2})) & (\textbf{M5}) \end{array} \tag{5}$$

Finally, for predicting the trading volume, we test:

$$\begin{array}{ll} \ln(\hat{v}_t) = f(\ln(v_{t-1})) & (\mathbf{M1}, \text{ baseline}) \\ \ln(\hat{v}_t) = f(\ln(v_{t-1}), \ln(\frac{n_{t-1}}{n_{t-2}})) & (\mathbf{M2}) \\ \ln(\hat{v}_t) = f(\ln(v_{t-1}), \ln(n_{t-1}), \ln(n_{t-2})) & (\mathbf{M3}) \\ \ln(\hat{v}_t) = f(\ln(v_{t-1}), \ln(MA_{t-1}), \ln(MA_{t-2})) & (\mathbf{M4}) \\ \ln(\hat{v}_t) = f(\ln(v_{t-1}), \ln(\frac{MA_{t-1}}{MA_{t-2}})) & (\mathbf{M5}) \end{array} \quad (6)$$

2.3 Evaluation

To measure the quality of predictions of the regression models, we use two error metrics, Root-Mean-Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), given by [5]:

$$\begin{array}{l} RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \\ MAPE = \frac{1}{N}\sum_{i=1}^{N}|\frac{y_i - \hat{y}_i}{y_i}| \times 100\% \end{array} \quad (7)$$

where y_i and $\hat{y}_i$ are the target and fitted value for the i-th day and N is the number of days considered. The lower the RMSE and MAPE values, the better the model. For the sake of correct evaluation, we apply both metrics, which can be used according to diverse goals. RMSE and MAPE compute the mean error but RMSE is more sensitive to high individual errors than MAPE (e.g. useful to avoid large error estimates).

For achieving a robust evaluation, we adopt a fixed-size (of length W) rolling windows evaluation scheme [15]. Under this scheme, a training window of W consecutive samples (last example corresponds to day $t-1$) is used to fit the model and then we perform one prediction (for day t). Next, we update the training window by removing the first observation in the sample and including the new observation for day t, in order fit the model and predict the value for $t+1$, an so on. For a dataset of length L, a total of $N = L - W$ predictions (and model trainings) are performed. In this paper, three different window sizes (i.e. $W \in \{100, 200, 300\}$) are explored.

We measure the value of StockTwits based data if the respective regression model is better than the baseline method. As a baseline method, we adopt an $AR(1)$ model. This regression model has only one input: the previous day (t-1) observation. To test the forecasting ability of the models, we apply the equality of prediction statistical test [4], under a pairwise comparison between the tested model and the baseline for MAPE and RMSE metrics.

3 Results

All experiments here reported were conducted using the open source **R** tool [11] and the programming language Python running on a Linux server. The StockTwits posts were delivered in JSON format, processed in **R** using the rjson package and stored using the MongoDB database format by adopting the

rmongodb package. The multiple regressions were run using the `lm` function of **R**.

The predictive errors for the returns are presented in Table 1. An analysis of this table reveals that the baseline is only outperformed in very few cases. Also, when better results are achieved, the differences tend to be small, specially for RMSE. In effect, there are only two cases where the differences are statistically significant (GS, $W = 300$; and SPX, $W = 300$), corresponding to the **M4** model and the MAPE metric.

Table 1. Returns predictive results (**bold** – better than baseline results, $\star$ – p-value < 0.05 and $\diamond$ – p-value < 0.10)

		RMSE					**MAPE (in %)**				
Stock	**W**	**M1**	**M2**	**M3**	**M4**	**M5**	**M1**	**M2**	**M3**	**M4**	**M5**
AAPL	100	0.0166	0.0167	**0.0165**	0.0166	0.0166	184	210	190	191	199
	200	0.0171	0.0172	0.0171	0.0171	0.0171	176	190	193	**174**	188
	300	0.0174	0.0174	0.0174	0.0174	0.0174	174	187	187	**171**	187
AMZN	100	0.0225	0.0228	0.0226	0.0225	0.0226	160	160	**156**	**155**	**153**
	200	0.0233	0.0236	0.0235	0.0233	0.0235	111	118	**109**	116	121
	300	0.0240	0.0243	0.0241	0.0241	0.0243	104	108	110	115	115
GOOG	100	0.0168	0.0169	0.0169	0.0171	0.0169	138	262	184	176	257
	200	0.0177	0.0179	0.0177	0.0177	0.0178	119	161	127	140	147
	300	0.0161	0.0161	**0.0160**	0.0161	0.0161	117	124	123	134	119
IBM	100	0.0124	0.0127	0.0124	0.0124	0.0126	137	139	147	145	**135**
	200	0.0127	0.0128	0.0127	0.0127	0.0128	145	**145**	**145**	154	**137**
	300	0.0130	0.0131	**0.0128**	**0.0128**	0.0130	129	**128**	137	145	**129**
GS	100	0.0213	0.0214	0.0213	**0.0212**	**0.0212**	129	146	140	131	141
	200	0.0225	0.0226	0.0225	0.0225	0.0225	119	135	124	**117**	124
	300	0.0234	0.0235	0.0235	0.0234	0.0234	116	124	**113**	**108***	**110**
SPX	100	0.0119	0.0121	0.0119	0.0119	0.0120	141	182	**141**	169	173
	200	0.0126	0.0126	**0.0125**	0.0126	0.0126	173	193	**159**	**166**	**168**
	300	0.0121	0.0121	**0.0119**	**0.0119**	**0.0119**	166	189	**138**	**142**$^{\diamond}$	**151**

Table 2 presents the forecasting results for volatility ($\ln(\sigma_t)$). For some stocks (e.g. AAPL and MAPE; AMZN and RMSE) there are several models that present lower errors when compared with the baseline. However, these differences are not statistically significant, since none of the tested models that include posting indicators outperformed the baseline.

The performances of the trading volume regressions is shown in Table 3. In contrast with the results obtained for the previous financial variables (returns and volatility), in this case there are some interesting results that are worth mentioning. There is a total of 16 models that statistically outperform the baseline under the RMSE metric. In particular, we highlight **M2**, which is statistically better than the baseline in 8 cases (corresponding to four stocks: AAPL, AMZN, GOOG and IBM), followed by **M3** (which outperforms the baseline in 5 cases, for AMZN and GOOG).

Table 2. Volatility predictive results (**bold** – better than baseline results)

		RMSE					MAPE (in %)				
Stock	**W**	**M1**	**M2**	**M3**	**M4**	**M5**	**M1**	**M2**	**M3**	**M4**	**M5**
AAPL	100	0.069	0.070	0.071	0.070	0.069	1.399	**1.397**	1.421	**1.392**	1.400
	200	0.070	0.070	0.070	0.070	**0.069**	1.389	**1.379**	1.389	**1.374**	1.391
	300	0.068	0.068	0.068	0.069	0.068	1.338	**1.334**	**1.337**	**1.335**	1.357
AMZN	100	0.062	**0.061**	**0.061**	**0.062**	**0.061**	1.090	1.121	1.121	1.119	1.135
	200	0.066	**0.064**	**0.063**	**0.065**	**0.063**	1.129	1.151	1.145	1.139	1.150
	300	0.067	**0.066**	**0.065**	**0.067**	**0.064**	1.132	1.142	1.140	**1.126**	1.144
GOOG	100	0.064	**0.064**	**0.064**	0.064	**0.064**	1.250	1.307	1.324	1.308	1.321
	200	0.065	**0.065**	**0.064**	0.065	**0.064**	1.230	1.318	1.326	1.279	1.301
	300	0.063	0.063	**0.063**	0.064	**0.063**	1.196	1.280	1.282	1.236	1.275
GS	100	0.070	**0.070**	**0.070**	**0.070**	**0.070**	1.323	1.334	1.336	1.335	1.338
	200	0.072	0.072	0.073	0.073	0.073	1.280	1.314	1.322	1.307	1.319
	300	0.065	0.065	0.065	0.066	0.065	1.171	1.212	1.208	1.212	1.207
IBM	100	0.067	0.067	0.067	0.068	**0.067**	1.458	1.475	1.471	1.479	1.514
	200	0.069	**0.069**	**0.069**	0.070	**0.068**	1.476	1.480	**1.471**	**1.471**	1.489
	300	0.055	**0.054**	**0.053**	**0.055**	**0.054**	1.248	**1.248**	**1.241**	**1.247**	1.267
SPX	100	0.074	**0.074**	**0.074**	**0.074**	0.075	1.787	1.789	1.797	1.804	1.811
	200	0.075	0.075	0.076	0.075	0.075	1.781	1.803	1.810	1.800	1.802
	300	0.069	0.070	0.070	0.069	0.069	1.695	1.734	1.736	1.723	1.725

Table 3. Trading volume predictive results (**bold** – better than baseline results, $\star$ – p-value < 0.05 and $\diamond$ – p-value < 0.10)

		RMSE					MAPE (%)				
Stock	**W**	**M1**	**M2**	**M3**	**M4**	**M5**	**M1**	**M2**	**M3**	**M4**	**M5**
AAPL	100	0.300	0.301	0.304	0.304	0.302	1.406	1.423	1.444	1.440	1.422
	200	0.279	**0.279***	0.280	0.281	0.280	1.334	1.340	1.358	1.355	1.340
	300	0.281	**0.281***	0.283	0.283	0.282	1.344	1.351	1.371	1.363	1.347
AMZN	100	0.322	**0.321**	0.324	0.322	**0.320**	1.599	1.616	1.631	1.645	1.635
	200	0.326	**0.324**$^\diamond$	**0.322***	**0.323**$^\diamond$	**0.325**$^\diamond$	1.631	1.647	1.636	1.656	1.659
	300	0.327	**0.324***	**0.324**$^\diamond$	0.328	**0.326**	1.633	1.644	1.647	1.676	1.659
GOOG	100	0.328	**0.325**$^\diamond$	**0.326***	0.328	0.329	1.662	1.669	1.693	1.684	1.674
	200	0.330	**0.323***	**0.323***	**0.325**$^\diamond$	**0.325**	1.675	**1.657**	1.680	1.687	**1.671**
	300	0.315	**0.308***	**0.311***	**0.312**	**0.310**	1.628	**1.605**	1.650	1.648	**1.615**
IBM	100	0.300	**0.299**	0.301	0.303	0.302	1.466	1.469	1.482	1.497	1.479
	200	0.309	**0.308**	**0.307**	**0.309**	0.309	1.557	1.564	1.562	1.558	1.568
	300	0.319	**0.316***	**0.317**	0.320	0.320	1.636	1.636	**1.636**	1.642	1.653
GS	100	0.323	0.324	0.325	0.329	0.327	1.568	1.573	1.577	1.596	1.590
	200	0.325	0.326	0.328	0.329	0.327	1.578	1.588	1.592	1.600	1.590
	300	0.321	0.322	0.322	0.323	0.321	1.583	1.585	1.588	1.601	1.588
SPX	100	0.203	0.204	0.206	0.205	0.205	0.623	0.634	0.630	0.626	0.629
	200	0.164	0.164	0.164	0.164	0.164	0.538	0.540	**0.535**	**0.532**	**0.536**
	300	0.160	0.160	0.161	0.160	0.160	0.526	0.528	0.531	**0.526**	**0.526**

For demonstration purposes, Figure 2 shows the quality of the fitted results for the best model for Google (**M2**, $W = 300$). The left of the plot shows the last fifty out-of-sample observations and predicted values (of a total of 305), revealing an interesting fit of the predictions. The right of the plot presents the scatter plot of observed versus predicted values (with all 305 predictions), showing the adequacy of the proposed linear model.

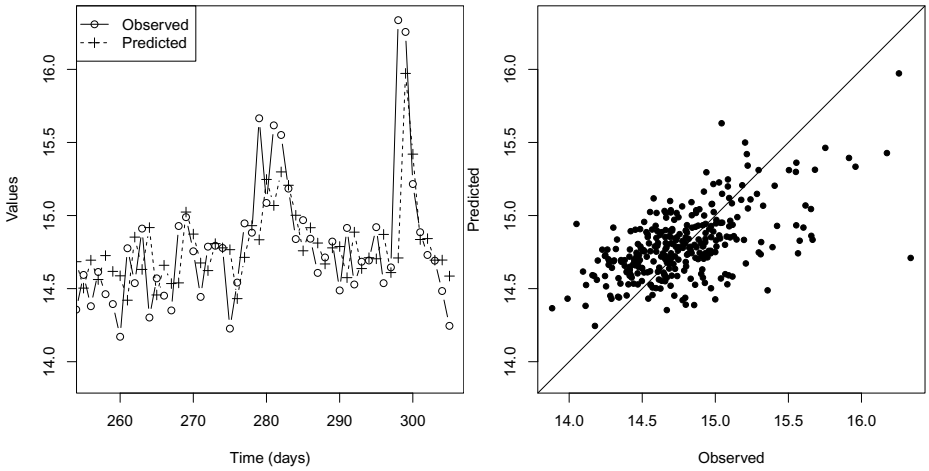

Fig. 2. Example of the last fifty trading volume ($\ln(v_t)$ out-of-sample values) and predictions for Google (GOOG), **M2** and $W = 300$ (left); and full scatter plot (with 305 predicted points) of observed versus predicted values (bottom, diagonal line denotes the perfect fit, natural logarithm values are used in both axis)

4 Conclusions

The main purpose of this study is to provide a more robust assessment about the relevance of microblogging data for forecasting three valuable stock market variables: returns, volatility and trading volume. We focused on a very recent and large dataset of messages collected from StockTwits, a social network service specifically targeted for communication about markets and their individual stocks. Following the recent related studies, we addressed two types of daily data from five large US companies and one major US index, sentiment indicators and posting volume, and explored four regression models that included input variables based on these microblogging indicators. However, and in contrast with previous works, we performed a more robust evaluation of the forecasting models, using fixed-sized rolling windows (with different window sizes), leading to much larger test periods (from 305 to 505 predictions). Also, we used two error metrics (MAPE and RMSE) and the equality of prediction statistical test, to compare the regression results with a baseline method (that uses only one input, the previous day stock market variable).

The results presented here suggest that predicting stock market variables using microblogging data, such as returns and volatility, is a much more complex and harder task than the previous and quite recent works presume. While this is an attractive research line, some caution is required when promising forecasting ability. This is not surprising considering the lack of support for return predictability in the Finance literature [16]. In this paper, we found scarce evidence for the utility of the tested sentiment variables when predicting returns, and of posting volume indicators when forecasting volatility. However, and aligned with the state of the art, we found interesting results when assessing the value of using posting volume for predicting trading volume, for two of the proposed regression models. We highlight that these results were obtained using the large test period so far, and much more stringent evaluation methods than previously.

In future work, we intend to explore more sophisticated parsers and lexicons, more adjusted to stock market terminology, to check if these can improve the investors' sentiment indicators. Also, we aim to address more complex regression models, using more time lags and complex base learners (e.g. Support Vector Machine [13]).

Acknowledgments. This work is funded by FEDER, through the program COMPETE and the Portuguese Foundation for Science and Technology (FCT), within the project FCOMP-01-0124-FEDER-022674. The also authors wish to thank StockTwits for kindly providing their data.

References

1. Amihud, Y., Mendelson, H., Pedersen, L.H.: Liquidity and Asset Prices. Foundations and Trends in Finance 1(4), 269–364 (2007)
2. Antweiler, W., Frank, M.Z.: Is all that talk just noise? the information content of internet stock message boards. The Journal of Finance 59(3), 1259–1294 (2004)
3. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science 2(1), 1–8 (2011)
4. Harvey, D., Leybourne, S., Newbold, P.: Testing the equality of prediction mean squared errors. International Journal of Forecasting 13(2), 281–291 (1997)
5. Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. International Journal of Forecasting 22(4), 679–688 (2006)
6. Jiang, G.J.: The Model-Free Implied Volatility and Its Information Content. Review of Financial Studies 18(4), 1305–1342 (2005)
7. Mao, H., Counts, S., Bollen, J.: Predicting financial markets: Comparing survey, news, twitter and search engine data. arXiv preprint arXiv:1112.1051 (2011)
8. Nofsinger, J.R.: Social mood and financial economics. The Journal of Behavioral Finance 6(3), 144–160 (2005)
9. Oh, C., Sheng, O.R.L.: Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. In: ICIS 2011 Proceedings (2011)
10. Peterson, R.L.: Affect and financial decision-making: How neuroscience can inform market participants. The Journal of Behavioral Finance 8(2), 70–78 (2007)
11. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2012) ISBN 3-900051-07-0

12. Schumaker, R.P., Chen, H.: Textual analysis of stock market prediction using breaking financial news: The azfin text system. ACM Transactions on Information Systems (TOIS) 27(2), 12 (2009)
13. Smola, A., Schölkopf, B.: A tutorial on support vector regression. Statistics and Computing 14, 199–222 (2004)
14. Sprenger, T., Welpe, I.: Tweets and trades: The information content of stock microblogs. Social Science Research Network Working Paper Series, pp. 1–89 (2010)
15. Tashman, L.J.: Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting 16(4), 437–450 (2000)
16. Timmermann, A.: Elusive return predictability. International Journal of Forecasting 24(1), 1–18 (2008)