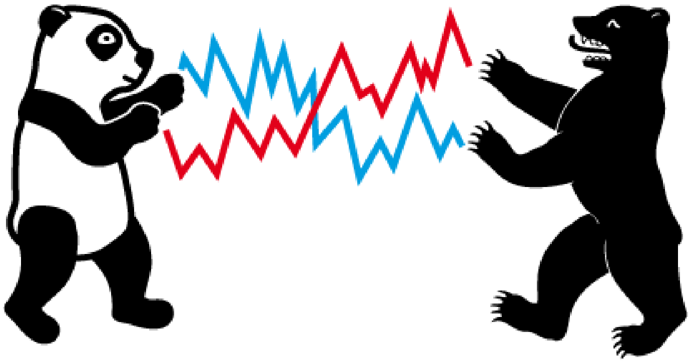
IRTG 1792 Discussion Paper 2018-066

Deep learning-based cryptocurrency sentiment construction

Sergey Nasekin [*]
Cathy Yi-Hsuan Chen [*2]



[*] Deutsche Bank AG, Germany
[*2] Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche
Forschungsgesellschaft through the
International Research Training Group 1792
"High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Deep learning-based cryptocurrency sentiment construction

Sergey Nasekin, [*] Cathy Yi-Hsuan Chen, [†]

December 10, 2018

Abstract

We study investor sentiment on a non-classical asset, cryptocurrencies using a "crypto-specific lexicon" recently proposed in Chen et al. (2018) and statistical learning methods. We account for context-specific information and word similarity by learning word embeddings via neural network-based Word2Vec model. On top of pre-trained word vectors, we apply popular machine learning methods such as recursive neural networks for sentence-level classification and sentiment index construction. We perform this analysis on a novel dataset of 1220K messages related to 425 cryptocurrencies posted on a microblogging platform StockTwits during the period between March 2013 and May 2018. The constructed sentiment indices are value-relevant in terms of its return and volatility predictability for the cryptocurrency market index.

Keywords: sentiment analysis, lexicon, social media, word embedding, deep learning

JEL Classification: G41, G4, G12

1 Introduction

The classical asset pricing theories, mainly relying on limit of arbitrage, meet challenges in the surge of brand new asset class like cryptocurrency. Compared to classical financial assets, the fundamental value, such as dividends, earnings and other type of cash flow, of new asset class are relatively intangible. The techniques behind cryptocurrency, such as blockchain, ICO (Initial Coin Offering), decentralized scheme, complicate the price evaluation, given the limited knowledge of investors.

Sentiment plays a role in the price evolution, given a possible arbitrage opportunity and intangible fundamental values, see Aboody et al. (2018). Cryptocurrency is exactly in this case. The

[*] Deutsche Bank AG, (*e-mail: sergey.nasekin@gmail.com*)

[†] C.A.S.E.- Center for Applied Statistics & Economics, Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin, Germany (*e-mail: cathy.chen@hu-berlin.de*)

question is how we measure the sentiment of it in a convincing way. The news about cryptocurrencies, among financial news for market or for stocks, account for a small proportion, due to its non-classical feature. In order to collect the representative sentiment/opinions from the crypto community who usually appear in social media, we target to StockTwits, a leading social media for financial discussion. This is motivated in Chen et al. (2018). They find that compared to Reddit offering the discussions focusing on crypto technologies, the sentiment distilled from StockTwits messages conveys more financial aspects, as a consequence, sentiment there gives better results on return predictability. In addition, microblogging users tend to react promptly to events, news and information, allowing a near real-time sentiment assessment.

We propose a state-of-art lexicon construction method used for cryptocurrency sentiment extraction, with automatic mining for large amounts of unstructured opinion content in order to summarize the opinions in the crypto community. Yet, the utilization of sentiment lexicons allows unsupervised classification of text, relieving the need for manual labeling of text. The existing lexicon hardly be employed for this task due to several reasons. Firstly, the domain-specific terms have been broadly employed to this community, e.g. "mining", "blockchain", "ICO", "wallet", "shitcoin", "binance", ”hodl”. Secondly, non-text characters that convey emotion, such as emojis and emoticons, appear very often in social media.

A pioneer study by Chen et al. (2018) shows the performance of cryptocurrency-specific sentiment, with a domain-specific lexicon created by the TF-IDF (Term Frequency - Inverse Document Frequency) scheme. However, the traditional bag-of-words approach does not account for context-specific dependencies between words and therefore important information about semantic structure of the sentence is lost. We therefore employ machine learning techniques such as Word2Vec introduced by Mikolov et al. (2013) and recurrent neural networks (RNN) to learn long-term semantic and syntactic dependencies in the messages.

We create crypto-sentiment indices by means of a predictive RNN model as well as utilizing an approach of lexicon expansion. We expand general social media lexica (viewed as seed lexica created by Renault (2017)) by incorporating domain-specific terms with a certain degree of similarity in their word embedding structures. The indices we create demonstrate predictive ability with respect to logarithmic returns of a cryptocurrency index "CRIX" developed by Trimborn and Härdle (2018) as well as its volatility. We expand standard predictive regression models for autoregressive mean and variance to show statistical significance of sentiment regressors.

2 Theoretical background

2.1 RNN architecture

Some of the most popular recurrent neural network (RNN) architectures applied for language modelling and sentiment prediction are the LSTM (long short-term memory) and GRU (gated recurrent unit) schemes, introduced by Hochreiter and Schmidhuber (1997) and Cho et al. (2014), respectively. A general architecture of a sentiment prediction LSTM/RNN network is presented in Figure 1. This architecture consists of the input sequence, an embedding lookup matrix, several layers of LSTM/GRU cells/units, an output sequence, mean pooling and softmax layers. The core of this structure are the LSTM or GRU cells. Structures of these cells are presented in Figures 2,

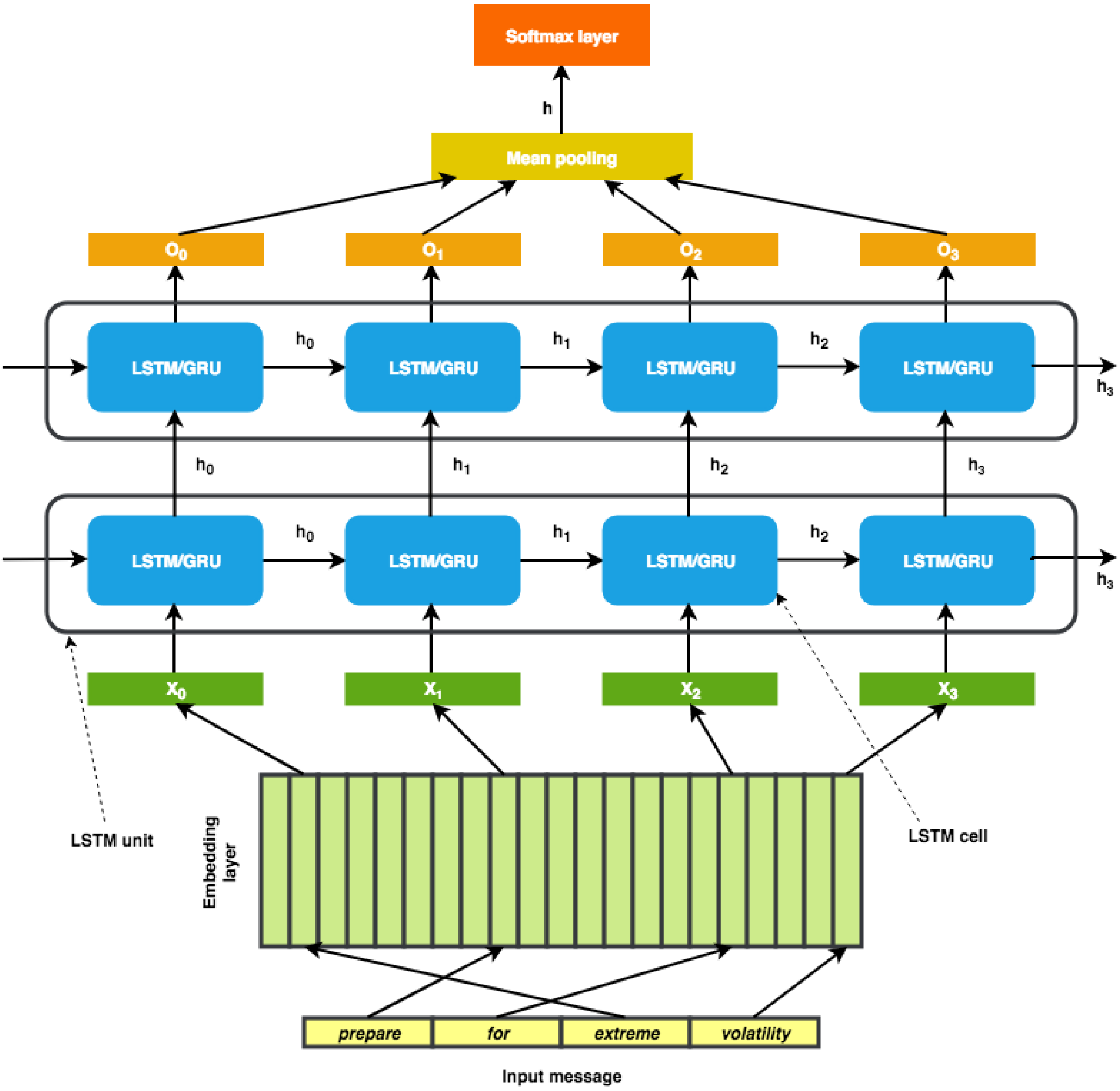


Figure 1: General architecture of an RNN

3, respectively.

The LSTM architecture in Figure 2 introduces the cell state C_t which is able to keep information about the previous states of LSTM cells. The amount of information stored in the cell state is controlled by the "gates": an input gate i_t, a forget gate f_t and an output gate g_t. The first to act is the forget gate f_t: it determines how much of the previous state C_{t-1} will be kept based on the values of the previous hidden state h_{t-1} and the current input x_t:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right), \tag{1}$$

where the sigmoid function $\sigma(x) = 1/(1+\exp(-x))$ outputs a value between 0 and 1 for each number in the cell state C_{t-1}.

Going further, the LSTM cell generates an update to C_{t-1} through a new candidate value of the

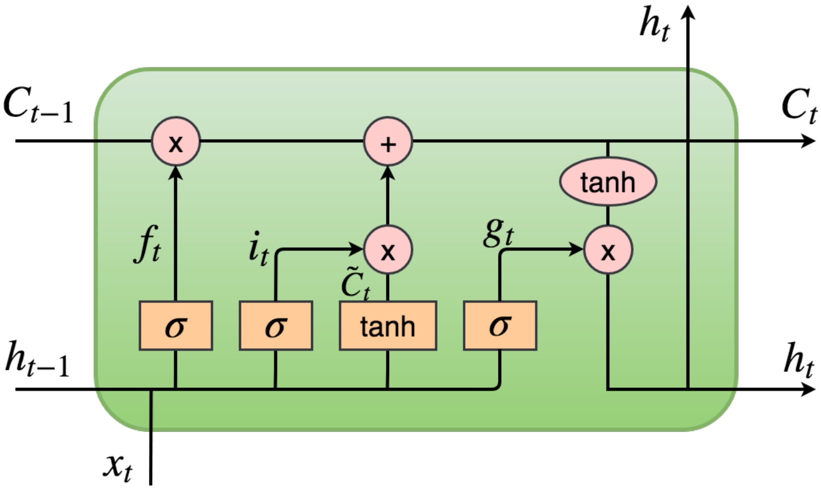


Figure 2: Structure of an LSTM unit

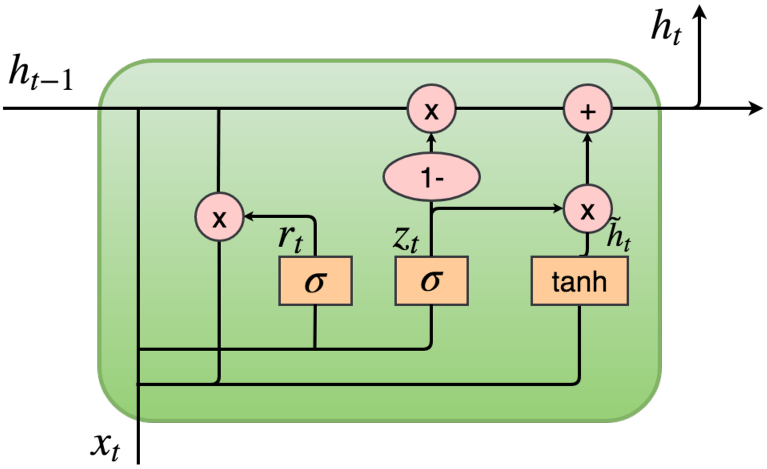


Figure 3: Structure of a GRU unit

cell state, $\tilde{C}_t$, which is created using a tanh layer:

$$\tilde{C}_t = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right), \tag{2}$$

where $\tanh(x) = \{\exp(x) - \exp(-x)\}/\{\exp(x) + \exp(-x)\}$.

Next, to decide what will be stored in the next cell state C_t, one has first to determine, "how much" of the new candidate state $\tilde{C}_t$ will be inputted into C_t. This is done through the input gate i_t, which, analogously to the situation with the forget gate f_t, outputs a number between 0 and 1 for each value of $\tilde{C}_t$:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right). \tag{3}$$

An updated value of the cell state C_t is essentially a weighted sum of the previous cell state value C_{t-1} and the new candidate value $\tilde{C}_t$:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \tag{4}$$

where $\odot$ denotes element-wise multiplication.

Finally, the next value of the hidden state h_t is a "filtered" value of the cell state C_t, which is put through the tanh nonlinearity and multiplied element-wise by the values of the output gate g_t:

$$h_t = g_t \odot \tanh(C_t), \tag{5}$$

where $g_t = \sigma(W_g x_t + U_g h_{t-1} + b_g)$. The resulting hidden state value h_t is propagated along LSTM cells within LSTM units as well as between the units and also upwards to the next hidden layer. Revisiting Figure 1, it should be noted that it is important to differentiate between LSTM cells and units. The former are the "black boxes" described by the equations above each of which works on a sequence element at time t. The latter are groups of LSTM cells.

The final output of a unit is a sequence $h_0, h_1, \ldots, h_n$ which is fed into the next unit as well as into the next layer. This type of complex architecture of deep LSTM networks allows them to manage long-term dependencies efficiently. This ability to balance "old" and "new" information through representations of recent input events yields the name "long short-term memory". Last but not least, this feature allows to mitigate the problem of vanishing gradients.

A similar system of equations describes the GRU architecture, which is a more parsimonious representation of a RNN unit similar to LSTM. In a GRU unit, as demonstrated in Figure 3, there is a "reset gate" r_t and an "update gate" z_t which combines forget and input of an LSTM unit. Furthermore, the cell state and hidden state are merged into one state h_t. The reset gate r_t determines how much of the past information contained in h_{t-1} is forgotten. The update gate z_t is a "decision rule" to construct a weighted average between the previous hidden state h_{t-1} and the new candidate value $\tilde{h}_t$. The system of equation describing this architecture is therefore as follows:

$$z_t = \sigma\left(W_z x_t + U_z h_{t-1}\right), \tag{6}$$

$$r_t = \sigma\left(W_r x_t + U_r h_{t-1}\right), \tag{7}$$

$$\tilde{h}_t = \tanh\left(W_h x_t + r_t \odot U_h h_{t-1}\right), \tag{8}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t. \tag{9}$$

2.2 Pre-trained embeddings

It has been observed that often the training algorithm converges more quickly and smoothly if the embedding matrix is pre-trained. In the literature, various methods to pre-train embedding weights have been proposed, among them models such as Word2Vec and GloVe.

The Word2Vec model is a family of methods including two methods for generating dense embeddings: skip-gram and CBOW (continuous bag of words). They are mirror images of each other: in skip-gram, one predicts the context words $c_1, c_2, \ldots, c_C$ from a given word w. In CBOW, it is vice versa: a word w is predicted from the context $c_1, c_2, \ldots, c_C$. Let us consider the more popular skip-gram model to realize how Word2Vec works: the goal is to maximize the conditional probability $p(c|w;\theta)$ of obtaining context words given the current word; this probability can be parameterized as a softmax:

$$p(c|w;\theta) = \frac{e^{v_c \cdot v_w}}{\sum\limits_{\substack{c' \in C \\ c' \neq c}} e^{v_{c'} \cdot v_w}}, \tag{10}$$

where v_c and $v_w \in \mathbb{R}^d$ are vector representations or *embeddings* of c and w, respectively. The parameters θ are v_{c_i}, v_{w_i} for $w \in P$, $c \in C$, where P and C are the vocabulary and the set of all contexts, respectively. The objective function to maximize is therefore

$$\arg\max_{\theta} \prod_{w \in P} \prod_{c \in C} p(c|w;\theta). \tag{11}$$

In fact, the objective in (11) is not practical as it is very computationally expensive to compute because of the summation over all c' in the denominator in (10). One popular solution to this problem is to use the so-called negative sampling which randomly samples several "noise" words from the corpus based on their frequency. This amounts to generating "normal" and "noise" pairs $(w, c) \in D$ and $(w, c) \in D'$, respectively, where $D \cup D'$ comprises the entire corpus.

Then the negative sampling objective can be obtained as follows:

$$\arg\max_{\theta} \sum_{(w,c) \in D} \log \sigma(v_c \cdot v_w) + \sum_{(w,c) \in D'} \log \sigma(-v_c \cdot v_w), \tag{12}$$

where σ is the softmax function. For more details, see Goldberg and Levy (2014).

The Word2Vec model is graphically represented as a shallow neural network model with one hidden layer with shared weights $\tilde{V}$ for all context words c; see Figure 4. In fact, the objective in (12) is just an approximation to the original objective (11) and as such does not produce optimal predictions for context words, but tends to produce meaningful embeddings V which can be further used in training a deep RNN model. Another useful feature of the Word2Vec approach is that it outputs dense vector representations for vocabulary words which have dimension d which can be much smaller than the size of the dictionary P.

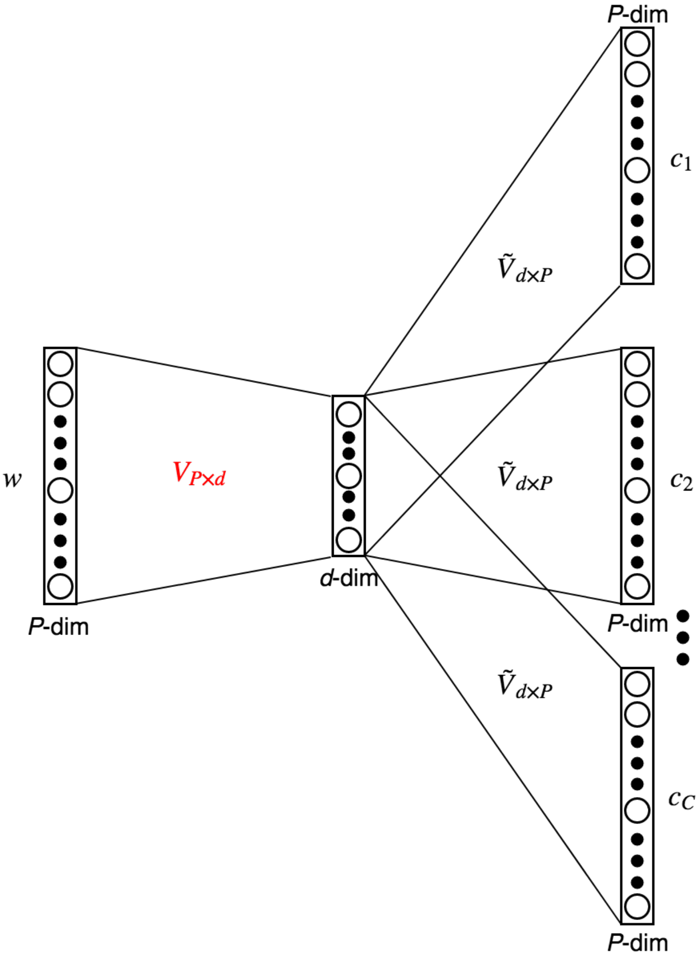


Figure 4: Structure of a Word2Vec neural model

3 Data and preprocessing

3.1 StockTwits

StockTwits[1], as a social microblogging platform, becomes popular and stands for a leading social network for investors and traders . In spite of its similarity to Twitter, the design of it is though oriented to financial discussion. One of features contributes to its popularity is an availability of messages and users streams related to financial assets (including cryptocurrencies), generated by investors or potential new comers. According to StockTwits, more than one million users now use the platform to share information and ideas, reaching an audience of more than 40 million people across the financial web and social media. Conversations are organized around "cashtags" (e.g. $SPY for S&P 500) that allows to narrow streams down to specific assets. Users can also express their sentiment by labeling their messages as "Bearish" (negative) or "Bullish" (positive) *via* a toggle button. As detailed by Chen et al. (2018) , the user generated messages and self-reported sentiment attract the researchers for sentiment analysis. The available labeled data benefits an advance on textual analysis that typically relies on the available training dataset.

Since 2014 StockTwits adds streams and symbology for cryptocurrencies and tokens, from 100+ in the beginning to 400+ recently. This brand new and vibrant new asset class have successfully attracted a huge attention from its big community and also from new comers. New cryptocurrencies are regularly added to the list of cashtags supported by StockTwits.[2] A cashtag refers to a cryptocurrency if and only if it ends with ".X" (e.g. $BTC.X for Bitcoin, $LTC.X for Litecoin). We use this convention and StockTwits Application Programming Interface (API) to download all messages containing a cashtag referring to a cryptocurrency. StockTwits API also provides for each message its user's unique identifier, the time it was posted at with a one-second precision, and the sentiment associated by the user ("Bullish", "Bearish" or unclassified). Our final dataset contains 1,220,728 messages from 33,613 distinct users, posted between March 2013 and May 2018, and related to 425 cryptocurrencies. Overall, 472,255 messages are classified as bullish (38.6%) and 92,033 as bearish (7.5%), and the remaining are unclassified. The imbalance between the numbers of positive and negative messages shows that online investors are optimistic on average, as previously found by Kim and Kim (2014) or Avery et al. (2016).

Figure 5 represents the number of messages per week related to cryptocurrencies on StockTwits, and CRIX (CRyptocurrency IndeX, see Trimborn and Härdle (2018)) weekly average. Investor attention has skyrocketed just like the prices did during the 2017 booming of the market, but it declines as the prices drop in 2018. This indicates a certain relationship between investors discussion on StockTwits and price movement.

3.2 Preprocessing

We follow the natural language processing implemented by Oliveira et al. (2016) and Chen et al. (2018). First, all messages are lower-cased. To collapse letter repetitions, which has been shown to be a critical feature of sentiment expression on microblogs (Brody and Diakopoulos; 2011),

[1] https://stocktwits.com/

[2] This list can be found at https://api.stocktwits.com/symbol-sync/symbols.csv.

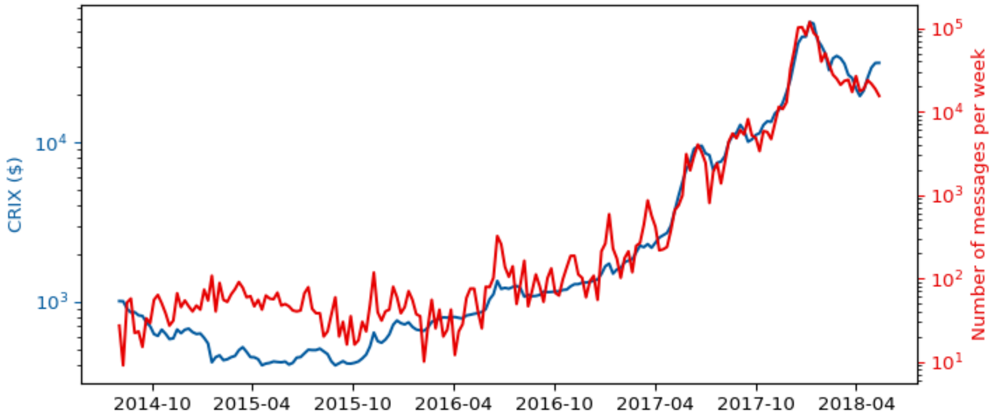


Figure 5: Weekly number of crypto-related messages on StockTwits and CRIX value (log scale).

Before processing	After processing
$BTC.X why can't it hold $14k ?? Shameless pumpers said 25 by Christmas "1F633	cashtag why can t it hold moneytag ? ? shameless pumpers said numbertag by christmas "1F633
$BTC.X Merry Xmas to all coiners and no coiners alike! 2018 is gonna be lit!! "1F4B0"1F384"1F680	cashtag merry xmas to all coiners and negtag_coiners alike ! numbertag is gonna be lit ! ! "1F4B0 "1F384 "1F680
$XVG.X all greeeeeeeeb "1F602"1F602	cashtag all greeeb "1F602 "1F602
$NEO.X In NEO I trust!!! https://neousd.bid/	cashtag in neo i trust ! ! ! linktag

Table 1: Pre-processing of StockTwits messages

sequences of repeated letters are shrinked to a maximum length of 3 (e.g. "Coooooool" will be truncated as "coool"). Tickers ("$BTC.X", "$ETH.X"...), dollar or euro values, hyperlinks, numbers and mentions of users are respectively replaced by the words "cashtag", "moneytag", "linktag", "numbertag" and "usertag". Substituting numbers by a single tag since the whole set of distinct numbers is too vast. For privacy reasons, all users and URL address are normalized to "usertag" and "linktage" respectively. We also exclude messages composed only by cashtags, URL links or simply punctuation. The prefix "negtag_" is added to any word consecutive to "not", "no", "none", "neither", "never" or "nobody".The stopwords and all punctuation except the characters "?" and "!" are removed. Exclamation and interrogation marks are kept as it has been previously shown that they are often part of significant bigrams that improve lexicon accuracy (Renault; 2017), which is also shown in Table 3 of Chen et al. (2018). Table 1 shows examples of messages before and after processing.

4 Textual analysis and sentiment prediction

4.1 RNN algorithm setup

Two features of StockTwits make itself popular not only from users' but also from researchers' perspective. First, messages contain explicit reference to the asset they mention *via* the "cashtag" system, which allows us to select messages that only refer to cryptocurrencies; second, users are encouraged to report their sentiment corresponding to the posted message, as "Bullish" (positive) or "Bearish" (negative), which rewards us a large training dataset adapted to supervised learning and for a cross-validation purpose.

We consider all messages labeled as "Bullish", and randomly split it into a training positive dataset (70% of all positive messages, i.e. 330,578) and a testing positive dataset (the remaining 30%, i.e. 141,676). To avoid domination of the corpus by excessively prolific users (possibly robots), we impose a maximum proportion of 1% of the dataset per user, as in Pang et al. (2002). Proceeding identically with negative messages, we constitute our final training and testing datasets.

The full filtered dataset of 528,443 messages is divided into the training subset of 422,754 messages (80% of all samples) and a test subset of 105,689 messages (20% of all samples). We use the technique of stratified sampling to ensure equal proportions of positive and negative messages in both train and test datasets.

Input data have an unbalanced structure: just about 16% of all labeled messages are bearish while the rest are bullish. Various methods have been proposed to address this problem:

- down-sampling the majority class,
- over-sampling the minority class,
- more advanced techniques such as SMOTE (Synthetic Minority Oversampling Technique) see Chawla et al. (2011).

Word2Vec			
Algorithm		Skip-gram	
Embedding dimension		256	
Context window size		10	
Number of epochs		25	
Common parameters			
Maximum encoded message length		50	
Unknown embedding vector		$U[-0.1, 0.1]$	
Loss function		Binary cross-entropy	
Batch size		64	
Number of epochs		50	
LSTM		**GRU**	
Recurrent layers	2	Recurrent layers	3
Recurrent unit	64	Recurrent layers	128
Recurrent dropout	50%	Recurrent layers	50%
Dropout	50%	Dropout	50%
Activation	tanh	Activation	tanh
Optimizer	Adadelta	Optimizer	Adadelta

Table 2: Parameters' setup for RNN

We apply oversampling by a factor of 5 to under-represented bearish messages.

We set up recurrent neural network models using methodology from Section 2, presented in Table 2. Two different approaches are tested: in one, embeddings are initialized randomly, in another, they are pre-trained using the Word2Vec model. We compare the performance of both setups and use the trained embeddings to construct sentiment indices.

4.2 Estimation results

Performance metrics of the tested RNN setups are shown in Table 3. These are the results of testing on the test set of 105,689 unique messages in total, without the count of over-sampled bearish messages. From the results above it follows that the LSTM deep RNN setup with word embeddings pre-trained by Word2Vec demonstrates better performance in terms of overall precision and accuracy than other setups. It has lower precision for bearish messages than the LSTM RNN setup with randomly initialized embeddings.

Lower precision for bearish messages of 43% and 50%, respectively, is caused by a higher rate of false positives when regarding bearish messages as positives as there are many more bullish messages ("negatives") which become false positives. Recall is not affected by this problem, however.

		Accuracy	Precision	Recall	F1-score	Data
LSTM (pre-trained embeddings)	Bullish	0.81	0.94	0.81	0.87	88,466
	Bearish	0.73	0.43	0.73	0.54	17,223
	Weighted avg. / Total	0.80	0.86	0.80	0.82	105,689
GRU (pre-trained embeddings)	Bullish	0.85	0.96	0.66	0.78	88,466
	Bearish	0.66	0.32	0.85	0.47	17,223
	Weighted avg. / Total	0.82	0.86	0.69	0.73	105,689
LSTM (random embeddings)	Bullish	0.89	0.92	0.89	0.90	88,466
	Bearish	0.58	0.50	0.58	0.54	17,223
	Weighted avg. / Total	0.84	0.85	0.84	0.84	105,689
GRU (random embeddings)	Bullish	0.78	0.93	0.78	0.85	88,466
	Bearish	0.69	0.38	0.69	0.49	17,223
	Weighted avg. / Total	0.77	0.84	0.77	0.79	105,689

Table 3: Performance metrics for RNN models

Other things being equal, we would prefer a higher false negatives' rate for bullish messages than a high false positives' rate for bearish samples as this might imply an underestimation of risk.

The deep LSTM model is able to capture similarities between words at least as well as the Word2Vec model. It identifies semantic connections between terms while using dynamically trained memory states rather than static context windows. This happens though at the cost of higher computational overhead.

Once the term embeddings have been obtained, we use a seed dictionary of 1311 terms from Renault (2017) constructed for StockTwits forums. The terms collected in this seed lexicon are commonly general across different assets discussed in this social media, and can be viewed as "finance domain-general" lexica. Each word in the lexicon has a sentiment score in the interval of $[-1, 1]$. Then we augment it with new context-specific terms as follows:

1. for each seed word d_S, find 2 most similar non-seed words d_{NS} using the trained embeddings, determined by non-negative cosine similarity value $CS(d_S, d_{NS})$ with $CS(a, b)$ defined as

$$CS(a,b) = \frac{\sum_{i=1}^{L} a_i b_i}{\sqrt{\sum_{i=1}^{L} a_i^2}\sqrt{\sum_{i=1}^{L} b_i^2}}, \tag{13}$$

where a, b are numeric vectors of dimension L,

2. assign a sentiment score $SW(d_{NS})$ to the non-seed word d calculated as

$$SW(d_{NS}) = CS(d_S, d_{NS}) \times SW(d_S), \tag{14}$$

with $SW(d_S)$ actually known from the seed lexicon.

The seed lexicon is augmented by further 1,286 terms counting 2,597 words in total. An illustration of these additional terms is presented in Figure 6, which context-specific bullish and bearish terms are grouped into "word clouds" reflecting term frequency. The larger is the font of the term in the "word cloud", the higher is the frequency. Context-specific terms emerge in the visualization and sometimes otherwise positive words acquire negative connotation. For instance, emoji symbols like ... and ... have been included into the list of positive terms. On the other hand, words like "high" and "China" which otherwise would carry positive or neutral context, have been classified as negative.

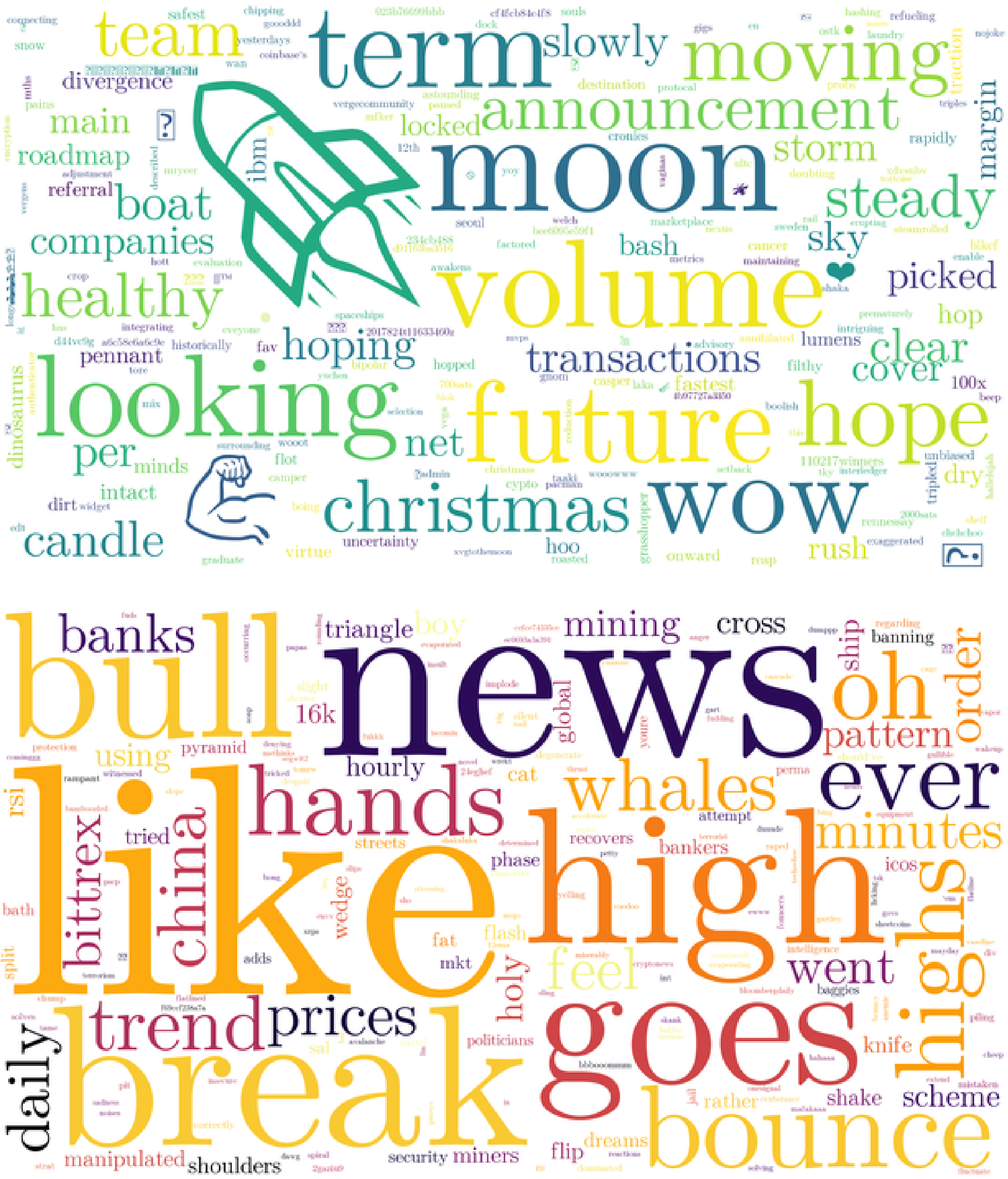


Figure 6: Word clouds for bullish (top) and bearish (bottom) context-specific terms augmenting the seed lexicon

4.3 Sentiment index construction

We create an aggregate cryptocurrency sentiment index quantified by the newly constructed domain-specific lexicon through an deployment of deep neural network for recursively learning domain-specific word embedding. This aggregate sentiment is deem a representative opinions from the crypto community in Stocktwits with their specific linguistic features. The information content of it is hypothetical to be relevant for future market performance and can be used to predict the price and volatility evolution, given the limited knowledge of fundamental value. Sentiment provides incremental explanatory power on firms' future performance, especially when fundamental information is incomplete or biased (see Tetlock et al. (2008); Lerman and Livnat (2010); Feldman et al. (2010); Loughran and McDonald (2011)).

The crypto-sentiment index is quantified by averaging the sentiment scores across the cryptocurrency-related message (with the cashtag ends with ".X" to indicate a crypto asset class). The sentiment score of individual message is calculated by averaging the sentiment weights of crypocurrency-specific terms within the message. In the case of no detected crypocurrency-specific terms, the score of this message will turn to zero. Mathematically, it can be derived through

$$\text{score}_m = \sum_{d_m=1}^{d_m=N_m} SW(d_m), \;\; \text{where} \;\; d_m = d_S \cup d_{NS} \tag{15}$$

where N_m stands for total number of words in message m and d_m stands for the term d in message m. $SW(d_m)$ is the sentiment weight of term, designated by our constructed lexicon in the range between -1 and 1. An aggregate investor-wise opinion index at daily frequency, using equal weight of scores across messages within the same trading day, is generated as follows:

$$\text{sent}_t^{expand} = \sum_{m=1}^{m=M_t} \text{score}_m / M_t \tag{16}$$

where M_t stands for total number of messages at time t. The indices are smoothed with 7-day (1 week) moving average values to iron out idiosyncratic jumps in the individual investors' opinion measures, and they are further standardized to be tractable in the statistic sense.

Another way to construct a sentiment index is to use a trained RNN model to predict sentiment labels of unlabeled messages which constitute about 60% of the StockTwits' messages' dataset. We use the LSTM setup with pre-trained Word2Vec embeddings for this purpose. Aggregated sentiment is constructed in the following way:

1. as a logarithmic rate of change of the number of bullish and bearish messages on a day t:

$$\text{sent}_t^{lstm1} = \log\left(\frac{M_t^{Bu} - M_t^{Be}}{M_{t-1}^{Bu} - M_{t-1}^{Be}}\right), \tag{17}$$

2. as an alternative "bullishness" measure proposed by Antweiler and Frank (2004)

$$\text{sent}_t^{lstm2} = \log\left(\frac{1 + M_t^{Bu}}{1 + M_t^{Be}}\right), \tag{18}$$

where M_t^{Bu} and M_t^{Be} is the number of bullish and bearish messages on day t, respectively. In the next section, we perform econometric analysis of predictability of a cryptocurrency index CRIX using three sentiment indices defined above.

5 Implications to cryptocurrency index

5.1 Return predictability

Figure 7 displays an interplay between the time series of crypto-sentiment index and the CRIX return over time. Their coherence in terms of time series dynamics motives us an investigation on return predictability of sentiment index.

To examine the predictability of sentiment, we consider the standard predictive regression model as:

$$r_{m,t+1} = \alpha + \beta \text{sent}_t + \phi Z_t + \epsilon_{t+1}, \tag{19}$$

where $r_{m,t+1}$ is the log return of CRIX, Z_t is a vector of alternative predictors encompassing the logarithm of message volume (MsgVol) and the moving average of $r_{m,t}$ (MA) suggested by Detzel et al. (2018). sent_t is one of sentiment measures defined in (17), (18) or (16). Our main interest is to test the significance of β, given the presence of the competing predictors. Table 4 reports the results of in-sample predictive regressions, with a sample period from Aug. 2014 to May 2018. Three sentiment measures quantified by three different types of fashions, sent^{lstm1}, sent^{lstm2}, sent^{expand} along with the one by using seed lexicon sent^{seed} as benchmark are separately incorporated into the regression for a task of market return prediction. By comparing their significance levels, we observe that the sentiment indices encompassing domain-specific information all have a higher predictability than the domain-general one by seed lexicon. Especially, the one using the expanded lexicon stands out. It confirms that the relevant information for return prediction mainly stems from domain-specific characteristics augmented by sentiment measures. To be more specific, Chen et al. (2018) discovers that the topics contributing to predictability are related to the financial aspects e.g. market activities and transactions, whereas discussions about the the technology (blockchain, mining, wallet) in Reddit are less informative in terms of the short-run prediction.[3] Their research confirms the value of investigating StockTwits, as it is designed for financial aspect discussions. The topics there ought to reflect investors' outlook from global and industrial perspectives, which are relatively decisive for future price movement.

With sentiment being considered, we find the explanatory power of the technical indicators proposed by Detzel et al. (2018) has vanished. In addition, the message volumes of StockTwits, as a proxy of market attention, cannot provide any incremental information.

5.2 Deterministic role in conditional volatility process

Sentiment extracted by microblogging users who discuss cryptocurrency-related topics may potentially trigger the volatility of underlying market. A burst of online discussions, if relevant, renders market fluctuation. To incorporate the distilled sentiment into the dynamics of variance process, we deploy sentiment acting as an exogenous variable in the context of GARCH framework.

The sentiment-driven conditional variance process is established in favor of an integrated GARCH

[3] Reddit is a generic message board, and not a message board only dedicated to financial markets, allowing us to capture a wider number of topics related to cryptocurrencies including discussions about cryptocurrency technologies and the blockchain.

Dependent Variable	$r_{m,t+1}$			
$sent^{lstm1}$	0.369			
	(0.017)			
$sent^{lstm2}$		0.356		
		(0.022)		
$sent^{expand}$			0.321	
			(0.006)	
$sent^{seed}$				0.266
				(0.120)
$MsgVol$	-0.036	-0.030	0.032	-0.006
	(0.528)	(0.586)	(0.460)	(0.900)
$MA1$	-0.255	-0.254	-0.253	-0.263
	(0.173)	(0.175)	(0.176)	(0.159)
$MA2$	0.212	0.213	0.201	0.213
	(0.422)	(0.419)	(0.444)	(0.419)
$MA3$	-0.191	-0.189	-0.189	-0.185
	(0.533)	(0.535)	(0.536)	(0.544)
R-$square$	0.788	0.760	0.926	0.627

Table 4: Control market microstructure impact

This table reports the return forecasting results with the control variables for the h-day moving average effect (MA(h)) and message volume (MsgVol). $sent^{lstm1}$, $sent^{lstm2}$, $sent^{expand}$ are defined in (17), (18) and (16), respectively. The p-values reported in parentheses are computed using Newey-West standard errors. The sample period is 2014-08-01 – 2018-05-15. The value of R-square is shown in percentage.

type (Engle and Bollerslev; 1986), given a non-stationary nature of cryptocurrency variance process. The property of covariance-stationary in the second moment is often violated in the case of cryptocurrency asset, due to the presence of permanent shocks. The integrated GARCH (IGARCH) model by Engle and Bollerslev (1986) is proposed for coping with highly persistent variance.

The specification of the IGARCH with the exogenous variable X_t is:

$$\begin{aligned} e_t &= y_t - \mathsf{E}_{t-1} y_t \qquad (20) \\ e_t &= Z_t \sigma_t, \quad Z_t \sim t(\nu) \\ \sigma^2_{t+1} &= \alpha e^2_t + (1-\alpha)\sigma^2_t + \theta X_t \qquad (21) \end{aligned}$$

where $0 < \alpha < 1$, E_{t-1} is the expectation operator conditional on $t-1$, σ^2_t represents the conditional variance of the process at time t, $t(\nu)$ refers to the zero-mean t distribution with ν degrees of freedom. The IGARCH(1,1) is chosen based on the BIC criteria. X_t here is the squared sentiment measure.

As clearly seen in Table 5, sentiment drives the conditional variance process only when sentiment measures accommodate the domain-specific information and knowledge. The domain-general sentiment play less role in this specific asset class. In other words, the fluctuation in this market is more attributed to the domain-specific information or sentiment.

Figure 8 displays a number of themes in the lifetime of cryptocurrency market index, through a visualization of its conditional volatility along with absolute return. The simulation for variance

Coefficients	Estimates	robust std	p value
$sent^{lstm1}$			
α	0.16754	0.02579	0.000
θ	0.00118	0.00049	0.019
ν	3.34760	0.19149	0.000
$sent^{lstm2}$			
α	0.16953	0.02585	0.000
θ	0.00081	0.00034	0.016
ν	3.28249	0.19194	0.000
$sent^{expand}$			
α	0.14926	0.01831	0.000
θ	0.00021	0.00009	0.027
ν	3.65245	0.21875	0.000
$sent^{seed}$			
α	0.12015	0.02177	0.000
θ	0.00421	0.00336	0.192
ν	3.71465	0.20182	0.000

Table 5: Estimated coefficients of IGARCH(1,1) model

The robust version of standard errors (robust std) are based on the method of White (1982).

dynamics in the specification of (21) exhibits a reconciliation between the model-specific volatility driven by sentiment (here we show the case of sent$_{expand}$) and the realized volatility (using absolute return). This market has experienced a number of huge fluctuations starting from 2017 until first quarter of 2018. The sentiment-driven volatility model is capable of capturing the actual fluctuations. Not surprisingly, Figure 7 manifests sentiment remarkably in the corresponding time frame.

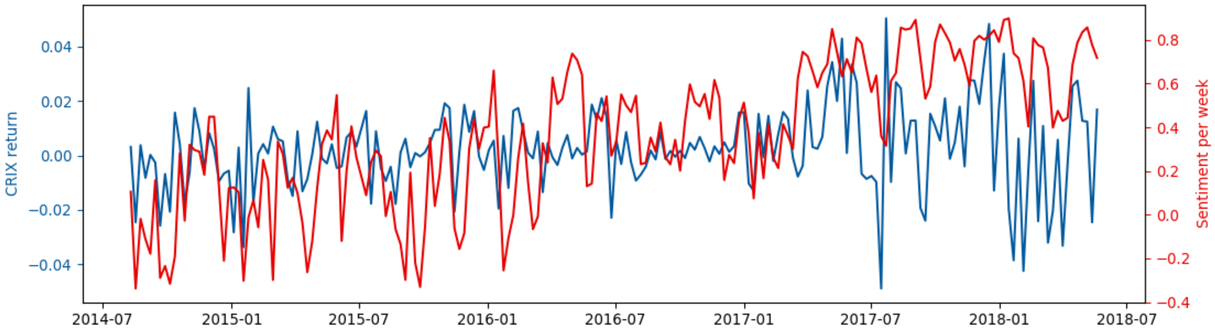


Figure 7: Co-movement between CRIX return and sentiment index (weekly basis)

The sentiment measure constructed through the LSTM method and defined in (17) is demonstrated.

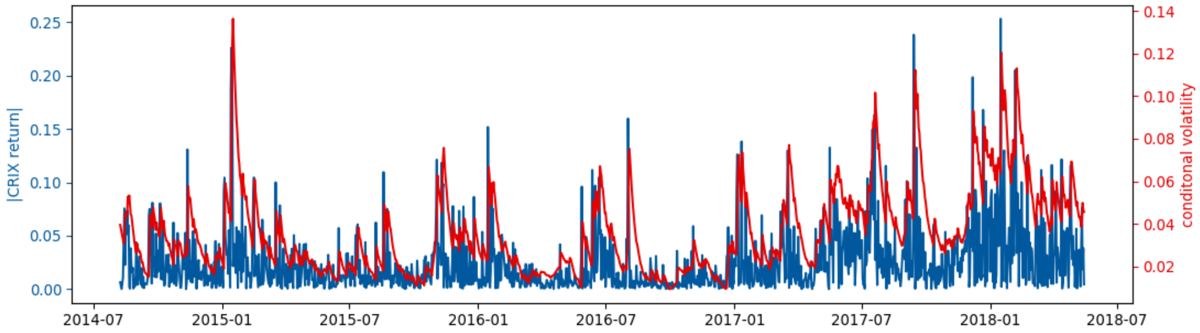


Figure 8: Sentiment-driven conditional volatility versus absolute return
The sentiment measure quantified by the expanded lexicon and defined in (16) is demonstrated.

6 Conclusion

In this paper, we study sentiments of cryptocurrency traders on StockTwits platform. We apply machine learning methods to construct sentiment indices reflecting opinions of cryptocurrency community on the market through time. Next, we integrate the newly built sentiment indices into predictive regressions for autoregressive mean and variance of cryptocurrency index' returns.

We observe that for an LSTM RNN setup, whether embeddings are pre-trained or not, does not play a significant role for the resulting performance of the predictive model. Nevertheless, the setup with pre-trained embeddings takes less time to train giving a decrease in computing overhead. Also this setup yields a more balanced outcome regarding individual performance for the bullish and bearish classes which potentially is more advantageous for under-represented bearish messages. Errors in prediction of bearish messages are more costly because they directly transform into under-estimation of downside risk.

We set up two types of predictive regressions for cryptocurrency index log-return time series: for the autoregressive mean and variance. In the first case, adding the constructed sentiment indices to the set of predictive variates significantly contributes to predictability of the log-returns. In the second setup, we find that there is presence of unit root in the GARCH specification of cryptocurrency returns' volatility. We therefore use an IGARCH approach with squared sentiment as an additional predictor. We find that sentiment contribution to crypto volatility prediction is significant. The sentiment-driven volatility model is capable of capturing the actual fluctuations of absolute returns of the cryptocurrency index.

References

Aboody, D., Even-Tov, O., Lehavy, R. and Trueman, B. (2018). Overnight returns and firm-specific investor sentiment, *Journal of Financial and Quantitative Analysis* **53**(2): 485–505.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards, *The Journal of Finance* **59**(3): 1259–1294.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2004.00662.x*

Avery, C. N., Chevalier, J. A. and Zeckhauser, R. J. (2016). The "caps" prediction system and stock market returns *, *Review of Finance* **20**(4): 1363–1381.
URL: *http://dx.doi.org/10.1093/rof/rfv043*

Brody, S. and Diakopoulos, N. (2011). Cooooooooooooooollllllllllllll!!!!!!!!!!!!!!: Using word lengthening to detect sentiment in microblogs, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 562–570.
URL: *http://dl.acm.org/citation.cfm?id=2145432.2145498*

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2011). Smote: Synthetic minority over-sampling technique.

Chen, C. Y., Després, R., Guo, L. and Renault, T. (2018). What makes cryptocurrencies special ? investor sentiment and price predictability in the absence of fundamental value, *Sfb 649 discussion paper*, Berlin.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation.
URL: *https://arxiv.org/abs/1406.1078*

Detzel, A. L., Liu, H., Strauss, J., Zhou, G. and Zhu, Y. (2018). Bitcoin: Learning, predictability and profitability via technical analysis.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances, *Econometric reviews* **5**(1): 1–50.

Feldman, R., Govindaraj, S., Livnat, J. and Segal, B. (2010). Management's tone change, post earnings announcement drift and accruals, *Review of Accounting Studies* **15**(4): 915–953.

Goldberg, Y. and Levy, O. (2014). Word2vec explained: deriving Mikolov et al.'s negative-sampling word-embedding method.
URL: *https://arxiv.org/abs/1402.3722*

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory, *Neural Computation* **9**(8): 1735–1780.

Kim, S.-H. and Kim, D. (2014). Investor sentiment from internet message postings and the predictability of stock returns, *Journal of Economic Behavior & Organization* **107**: 708 – 729. Empirical Behavioral Finance.
URL: *http://www.sciencedirect.com/science/article/pii/S0167268114001206*

Lerman, A. and Livnat, J. (2010). The new form 8-k disclosures, *Review of Accounting Studies* **15**(4): 752–778.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *The Journal of Finance* **66**(1): 35–65.
URL: *http://dx.doi.org/10.1111/j.1540-6261.2010.01625.x*

Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013). Efficient estimation of word representations in vector space.
URL: *https://arxiv.org/abs/1301.3781*

Oliveira, N., Cortez, P. and Areal, N. (2016). Stock market sentiment lexicon acquisition using microblogging data and statistical measures, *Decision Support Systems* **85**: 62 – 73.
URL: *http://www.sciencedirect.com/science/article/pii/S0167923616300240*

Pang, B., Lee, L. and Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques, *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 79–86.
URL: *https://doi.org/10.3115/1118693.1118704*

Renault, T. (2017). Intraday online investor sentiment and return patterns in the u.s. stock market, *Journal of Banking & Finance* **84**: 25 – 40.
URL: *http://www.sciencedirect.com/science/article/pii/S0378426617301589*

Tetlock, P. C., Saar-Tsechansky, M. and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals, *The Journal of Finance* **63**(3): 1437–1467.

Trimborn, S. and Härdle, W. K. (2018). Crix an index for cryptocurrencies, *Journal of Empirical Finance* **49**: 107–122.

White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica: Journal of the Econometric Society* pp. 1–25.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Data Driven Value -at -Risk Forecasting using a SVR -GARCH -KDE Hybrid" by Marius Lux, Wolfgang Karl Härdle, Stefan Lessmann, January 2018

002 "Nonparametric Variab le Selection and Its Application to Additive Models" by Zheng-Hui Feng, Lu Lin, Ruo -Qing Zhu as nd Li -Xing Zhu, January 2018

003 "Systemic Risk in Global Volatility Spillover Networks: Evidence from Option -implied Volatility Indices" by Zihui Yang, Yinggang Zhou, January 2018

004 "Pricing Cryptocurrency options: the case of CRIX and Bitcoin" by Cathy YH Chen, Wolfgang Karl Härdle, Ai Jun Hou, Weining Wang, January 2018

005 "Testing for b ubbles in cryptocurrencies with time -varying volatility" by Christian M. Hafner, January 2018

006 "A Note on Cryptocurrencies and Currency Competition" by Anna Almosova, January 2018

007 "Knowing me, knowing you: inventor mobility and the formation of technology -oriented alliances" by Stefan Wagner, Martin C. Go ossen, February 2018

008 "A Monetary Model of Blockchain" by Anna Almosova, February 2018

009 "Deregulated day -ahead electricity markets in Southeast Europe: Price forecasting and comparative structural analysis" by Antanina Hryshchuk, Stefan Lessmann, February 2018

010 "How Sensitive are Tail -related Risk Measures in a Contamination Neighbourh ood?" by Wolfgang Karl Härdle, Chengxiu Ling, February 2018

011 "How to Measure a Performance of a Collaborative Research Centre" by Alona Zharova, Janine Tellinger-Rice, Wolfgang Karl Härdle, February 2018

012 "Targeting customers for profit: An ensemble learning framework to support marketing decision making" by Stefan Lessmann, Kristof Coussement, Koen W. De Bock, Johannes Haupt, February 2018

013 "Improving C rime Count Forecasts Using Twitter and Taxi Data" by Lara Vomfell, Wolfgang Karl Härdle, Stefan Lessmann, February 2018

014 "Price Discovery on Bitcoin Markets" by Paolo Pagnottoni, Dirk G. Baur, Thomas Dimpfl, March 2018

015 "Bitcoin is not the New Gold - A Comparison of Volatility, Correlation, and Portfolio Performance" by Tony Klein, Hien Pham Thu, Thomas Walther, March 2018

016 "Time -varying Limit Order Book Networks" by Wolfgang Karl Härdle, Shi Chen, Chong Liang, Melanie Schienle, April 2018

017 "Regularization Approach for NetworkModeling of German EnergyMarket" by Shi Chen, Wolfgang Karl Härdle, Brenda López Cabrera, May 2018

018 "Adaptive Nonparametric Clustering" by Kirill Efimov, Larisa Adamyan, Vladimir Spokoiny, May 2018

019 "Lasso, knock off and Ga ussian covariates: a comparison" by Laurie Davies, May 2018

020 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin, Yanli Zhu, May 2018

021 "LASSO -Driven Inference in Time and Space" by Victor Chernozhukov, Wolfgang K. Härdle, Chen Huang, Weining Wang, June 2018

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

022 "Learning from Errors: The case of monetary and fiscal policy regimes" by Andreas Tryphonides, June 2018

023 "Textual Sentiment , Option Characteristics, and Stock Return Predictability" by Cathy Yi-Hsuan Chen, Matthias R. Fengler, Wolfgang Karl Härdle, Yanchu Liu, June 2018

024 "Bootstrap Confidence Sets For Spectral Projectors Of Sample Covariance" by A. Naumov, V. Spokoiny, V. Ulyanov, June 2018

025 "Construction of Non -asymptotic Confidence Sets in 2 -Wasserstein Space" by Johannes Ebert, Vladimir Spokoiny, Alexandra Suvorikova, June 2018

026 "Large ball pro babilities, Gaussian comparison and anti -concentration" by Friedrich Götze, Alexey Naumov, Vladimir Spokoiny, Vladimir Ulyanov, June 2018

027 "Bayesian inference for spectral projectors of covariance matrix" by Igor Silin, Vladimir Spokoiny, June 2018

028 "Toolbox: Gaussian comparison on Eucledian balls" by Andzhey Koziuk, Vladimir Spokoiny, June 2018

029 "Pointwise adaptation via stagewise ag gregation of local estimates for multiclass classification" by Nikita Puchkin, Vladimir Spokoiny, June 2018

030 "Gaussian Process Forecast with multidimensional distributional entri es" by Francois Bachoc, Alexandra Suvorikova, Jean-Michel Loubes, Vladimir Spokoiny, June 2018

031 "Instrumental variables regression" by Andzhey Koziuk, Vladimir Spokoiny, June 2018

032 "Understanding Latent Group Structure of Cryptocurrencies Market: A Dynamic Network Perspective" by Li Guo, Yubo Tao, Wolfgang Karl Härdle, July 2018

033 "Optimal contracts under competition when uncertainty from adverse selection and moral hazard are present" by Natalie Packham, August 2018

034 "A factor -model appro ach for correlation scenarios and correlation stress -testing" by Natalie Packham, Fabian Woebbeking, August 2018

035 "Correlation Under Stress In Normal Variance Mixture Models" by Michael Kalkbrener, Natalie Packham, August 2018

036 "Model risk of contingent claims" by Nils Detering, Natalie Packham, August 2018

037 "Default probabilities and default correlations under stress" by Natalie Packham, Michael Kalkbrener, Ludger Overbeck, August 2018

038 "Tail -Risk Protection Trading Strategies" by Natalie Packham, Jochen Papenbrock, Peter Schwendner, Fabian Woebbeking, August 2018

039 "Penalized Adaptive Forecasting with Large Information Sets and Structural Changes" by Lenka Zbonakova, Xinjue Li, Wolfgang Karl Härdle, August 2018

040 "Complete Convergenc e and Complete Moment Convergence for Maximal Weighted Sums of Extended Negatively Dependent Random Variables" by Ji Gao YAN, August 2018

041 "On complete convergence in Marcinkiewicz -Zygmund type SLLN for random variables" by Anna Kuczmaszewska, Ji Gao YAN, August 2018

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche
Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

042 "On Complete Convergence in Marcinkiewicz -Zygmund Type SLLN for END Random Variables and its Applications" by Ji Gao YAN, August 2018

043 "Textual Sentiment and Sector specific reaction" by Elisabeth Bommes, Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle, September 2018

044 "Understanding Cryptocurrencies" by Wolfgang Karl Härdle, Campbell R. Harvey, Raphael C. G. Reule, September 2018

045 "Predicative Ability of Similarity -based Futures Trading Strategies" by Hsin-Yu Chiu, Mi-Hsiu Chiang, Wei-Yu Kuo, September 2018

046 "Forecasting the Term Structure of Option Implied Volatility: The Power of an Adaptive Method" by Ying Chen, Qian Han, Linlin Niu, September 2018

047 "Inferences for a Partially Varying Coefficient Model With Endoge nous Regressors" by Zongwu Cai, Ying Fang, Ming Lin, Jia Su, October 2018

048 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin, Yanli Zhu, October 2018

049 "Strict Stationarity Testing and GLAD Estimat ion of Double Autoregressive Models" by Shaojun Guo, Dong Li, Muyi Li, October 2018

050 "Variable selection and direction estimation for single -index models via DC -TGDR method" by Wei Zhong, Xi Liu, Shuangge Ma, October 2018

051 "Property Investment and Rental Rate under Housing Price Uncertainty: A Real Options Approach" by Honglin Wang, Fan Yu, Yinggang Zhou, October 2018

052 "Nonparametric Additive Instrumental Variable Estimator: A Group Shrinkage Estimation Perspective" by Qingliang Fan, Wei Zhong, October 2018

053 "The impact of temperature on gaming productivity: evidence from online games" by Xiaojia Bao, Qingliang Fan, October 2018

054 "Topic Modeling for Analyzing Open -Ended Survey Responses" by Andra-Selina Pietsch, Stefan Lessmann, October 2018

055 "Estimation of the discontinuous leverage effect: Evidence from the NASDAQ order book" by Markus Bibinger, Christopher Neely, Lars Winkelmann, October 2018

056 "Cryptocurrencies, Metcalfe's law and LPPL models" by Daniel Traian Pele, Miruna Mazurencu-Marinescu-Pele, October 2018

057 "Trending Mixture Copula Models with Copula Selection" by Bingduo Yang, Zongwu Cai, Christian M. Hafner, Guannan Liu, October 2018

058 "Investing with cryptocurrencies Œ evaluating the potential of portfolio allocation strategies" by Alla Petukhina, Simon Trimborn, Wolfgang Karl Härdle, Hermann Elendner, October 2018

059 "Towards the interpretation of time -varying regularization parameters in streaming penalized regression models" by Lenka Zbonakova, Ricardo Pio Monti, Wolfgang Karl Härdle, October 2018

060 "Residual's Influence Index (Rinfin), Bad Leverage And Unmaskin g In High Dimensional L2 -Regression" by Yannis G. Yatracos, October 2018

061 "Plug -In L2 -Upper Error Bounds In Deconvolution, For A Mixing Density Estimate In R d And For Its Derivatives" by Yannis G. Yatracos, October 2018

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

062 ”Conversion uplift in e -com merce: A systematic benchmark of modeling strategies” by Robin Gubela, Artem Bequé, Fabian Gebert, Stefan Lessmann, November 2018

063 ”Causal I nference using Machine Learning. An Evaluation of recent Methods through Simulations” by Daniel Jacob, Stefan Lessmann, Wolfgang Karl Härdle, November 2018

064 ”Semiparametric Estimation and Variable Selection for Single-index Copula Models” by Bingduo Yang, Christian M. Hafner, Guannan Liu, Wei Long, December 2018

065 ”Price Management in the Used-Car Market: An Evaluation of Survival Analysi” by Alexander Born, Nikoleta Kovachka, Stefan Lessmann, Hsin-Vonn Seow, December 2018

066 ”Deep learning-based cryptocurrency sentiment construction” by Sergey Nasekin, Cathy Yi-Hsuan Chen, December 2018

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche
Forschungsgemeinschaft through the IRTG 1792.