Proceedings Paper:

Qu, Haizhou and Kazakov, Dimitar Lubomirov orcid.org/0000-0002-0637-8106 (Accepted: 2019) Detecting Causal Links between Financial News and Stocks. In: Proceedings of IEEE Conference on Computational Intelligence for Financial Engineering and Economics: (CIFEr 2019). 2019 IEEE Conference on Computational Intelligence for Financial Engineering and Economics, 04 May 2019 , CHN . (In Press)

Detecting Causal Links between Financial News and Stocks

Haizhou Qu
Department of Computer Science, University of York
York, United Kingdom, YO105GH
hq524@york.ac.uk

Dimitar Kazakov
Department of Computer Science, University of York
York, United Kingdom, YO105GH
dimitar.kazakov@york.ac.uk

***Abstract*—This article describes a novel framework for the detection of causal links between financial news and the subsequent movements of the stock market. The approach builds on and substantially improves a previously published in-house design for the detection and measurement of correlation between news and time series in the financial domain, which has been used here to detect a predictive causality relationship from news to prices and volumes of trade. While the original framework makes use of matrices of pairwise distances between companies, one based on news, the other – on financial performance, in order to produce a single measure of correlation between these two types of information for all traded companies, this article shows how the company contributing the most to the news-to-price/volume causal link can be singled out. The potential benefits of such information are made clear through its use in a straight-forward trading strategy, the results of which compare favourably to two strong, real-life alternatives that only make use of the time series.**

I. Introduction

The assumption that financial news can and should inform trading decisions may appear entirely obvious, yet the reality is more nuanced. Automated, high frequency trading (HFT) algorithms make decisions in fractions of a second based on numerical data alone. A popular tenet of economic theory, the efficient market hypothesis [1], which we have questioned in the past [2], effectively treats the impact of any relevant information on the stock price as an instantaneous process in which trades correct the price to the point where the information is no longer useful. The question this article addresses is whether it is possible to find evidence of news *driving* the markets, rather than providing a running commentary of present and past stock movements. While the novel method proposed here is universal, the data used in the experiments only makes it possible to discover such a causal link on the time scale of several days. On one hand, this is a strong constraint, which makes our task harder. On the other hand, being able to establish that news about company X can be used to trade its stock on the time scale of days, rather than a split second, would mean that (1) it is still viable to take the additional time needed to collect and process the news, and that (2) the results can be used by human traders, rather than algorithms alone. By implication, our approach can be used by 'ordinary' investors who can take their time, rather than be limited to, say, *market makers* with their exclusive advantage of receiving information on the buy/sell orders a few miliseconds early or traders with company servers placed in the immediate vicinity of a stock exchange in order to minimise the time the signals take to travel.

We have already established that for a given set of companies, the financial news content and stock market time series observed over the same time period can show substantial, statistically significant levels of correlation, which vary with time [3]. The approach we employed to demonstrate this made use of two matrices of pairwise distances between those companies, one based on news, the other—on financial performance (price or volume of trade). We used the Mantel test [4] to produce a single measure of correlation between these two matrices, and the two types of information they represent, for all traded companies.

In more detail, the above-mentioned approach requires a number of design choices to be made. Here we list all alternatives studied in our previously published work [3].

1) Firstly, one needs to choose a way to represent all news articles about a given company published in a given time period. We have experimented with (a) the TF-IDF weighted bag-of-words (BoW) representation [5], and (b) word embeddings (*aka* Word2Vec) [6].
2) Secondly, a distance measure between two sets of news items needs to be selected. For either representation, these were mapped on to a pair of vectors, which we compared through the cosine distance (CD) or Euclidian distance (ED).
3) Thirdly, a distance measure between two time series needs to be selected. Since each series is a vector, CD and ED are again applicable. In addition, we also tested Pearson's correlation as another distance metric for this type of data.

In this article we study several new alternatives to the above choices. The word mover's distance (WMD) is used as a text metric, while dynamic time warping (DTW) is used to compare two time series, adding the benefit of being also applicable to time series of different length.

The Mantel test is limited to measuring linear correlation. Therefore, we have also added Spearman's Rank-Order correlation as a way of detecting non-linear correlation between news and time series. In either case, we have sampled the data using non-overlapping windows, so that we measured the correlation between the financial news over a given period and the price or volume of stock traded over a subsequent time period. Such correlation can be interpreted as an indicator of *predictive causality* [7], as present news can be used to forecast how stock will trade in the future.

If no substantial causal link is detected by either test, this would imply that the news contained in the selected time window has nothing to contribute to the time series forecast, at least on the time scale represented in the data. However, the experimental results in this paper show that such a link exists for a significant proportion of the time period studied (with Spearman's correlation outperforming the Mantel test). This in itself is a new, important and non-trivial result: in simple words, it means that one can detect when financial news from the last few days has the potential to help one's choice of daily trade, and that such opportunities are frequent.

Nevertheless, the detection of a causal link between news and price (or volume of trade) for a *set of companies* is not easy to monetise, as in its basic form it does not suggest any particular trade. Another important contribution of this paper is that we show how this result can be turned into actionable information. The idea is straight-forward: for a set of N companies, run the test as many times, each time leaving the news and time series data on one company out, then single out the company whose absence from the dataset leads to the largest relative weakening of the detected causal link. This result means that the news about the company in question contains the largest amount of information with a bearing on its financial performance.[1] This can be used to draw the attention of a human trader, who can then study the highlighted news and draw their own decision whether to buy or sell the stock in question. It is possible that this step could be automated, e.g. through the use of sentiment analysis [8]. While the necessary techniques are well studied, and off-the-shelf tools exist, we have not experimented with them yet.

We have however tested the benefits of a trading strategy that assumes the company stock singled out by our approach would always go up on the following day. This, of course, may not be the case, and we ourselves have discussed an extreme counter-example where the release of negative news over several days appeared to produce a persistent downward trend in the affected stock price [9]. Still, experimenting with this strategy provides us with a conservative estimate of the full potential of our approach, and for the period of bullish market represented in the data, it compares favourably with two strong alternative strategies.

The rest of the article is structured as follows. Section II, describes the dataset used in the paper, and the ways in which it was preprocessed and sampled. Section III describes the distance metrics applied to news and to time series, the correlation tests used, and how the leave-one-out strategy was implemented. Section IV describes the design and results of our experiments, while Section V draws conclusions and discusses future work.

[1]To be more precise, news about that company differs from the news about all other companies in a way that most closely matches the differences between the company's subsequent financial performance and the achievements of all other companies.

TABLE I: Data Overview

Item	**Value**
Date	2014-10-1 to 2015-4-30
Number of stocks	25
Number of news items	189,151
Number of calendar days	212
Number of trading days	142

II. Data

We collected online news from Yahoo over the period 1 Oct 2014 – 30 Apr 2015. Each news article carries an EST time stamp and the symbol of one or more stocks, to which it is related. 25 stocks were selected on the basis of having no more than 5 calendar days with no news about them in the studied period. Very short news items of less than 10 words or less than 100 characters were ignored, leaving a total of 189,151 news items. OHLCV stock trading data (that is, Open, High, Low and Closing price, and Volume of trade) was collected from Yahoo Finance for the same period. Tables I and III provide an overview of all data used here. We have also generated additional time series following the equations below:

$$r_t = \frac{o_t - p_t}{o_t} \tag{1}$$

$$r'_t = \frac{a_t - a_{t-1}}{a_{t-1}} \tag{2}$$

$$c_t = \frac{h_t - l_t}{o_t} \tag{3}$$

where o, h, l and p are the time series of open, high, low and closing price respectively; a is the *adjusted closing price*[2] (ADJ); r is the *intraday return* (IRT) measuring the intraday profit ratio gained by a stock; r' stands for the *overnight return* (ORT), which measures the overnight profit ratio of a stock, and c is the *day change* (CHG), representing a metric of volatility. We also use the abbreviation VOL to refer to the time series representing the volume of trade.

III. Method

A. Measuring Correlation between News and Time Series

1) The Mantel Test: This is a statistical test defined by Equation 4, which is used to measure the linear correlation r between two $n \times n$ pairwise distance matrices $\mathbf{U}$ and $\mathbf{V}$ [4], e.g. to find out how much the geographic distance between two species is correlated with their genetic differences.

$$r_{\mathbf{U},\mathbf{V}} = \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{U_{ij} - \bar{u}}{\sigma_{\mathbf{U}}} \cdot \frac{V_{ij} - \bar{v}}{\sigma_{\mathbf{V}}} \tag{4}$$

Here $m = n(n-1)/2$ is the number of pairwise distances of a population of size n, and $\bar{u}$, resp. $\bar{v}$ are the means of the

[2]An adjusted closing price is a stock's closing price amended to the price taking dividends, splits, right offers etc. into account, so that the price is comparable with previous closing prices.

pairwise distance elements located above the main diagonal of $\mathbf{U}$, resp. $\mathbf{V}$.

2) Spearman's Rank Correlation Coefficient: This is a nonparametric test that measures the strength and direction of association between two variables. Given n pairs of observations, Spearman's rank correlation coefficient r_S can be calculated using the formula for the Pearson correlation coefficient in which the value of each observation is replaced with its rank (for the given variable). As a result, nonlinear correlations can also be detected and quantified.

$$r_S = \frac{cov(rg_U, rg_V)}{\sigma_{rg_U}\sigma_{rg_V}} \tag{5}$$

where rg_U and rg_V are the rankings of variables U and V, $cov(rg_U, rg_V)$ is their covariance, and σ_{rg_U} and σ_{rg_V} are their standard deviations.

To apply this test to the two matrices of distances, they are first converted into two paired variables by pairing the text distance between each two companies with the time series distance for the same two companies. The test is then applied in the usual way.

As already mentioned in the introduction, we use either test to determine the extent to which financial news and stocks are correlated with each other. The Mantel test only measures linear correlation. Considering the disparate nature of the two types of distances studied, it appears wise to assume that any potential dependence may be of nonlinear nature, which is why Spearman's rank correlation coefficient has been added to the list of options.

B. Measuring Distance between Time Series

Our previously published research assumed time series of equal length and experimented with three distance metrics: cosine distance (CD), Euclidean distance (ED) and Pearson's correlation (PD), which were defined as follows:

$$CD(\mathbf{x},\mathbf{y}) = 1 - \frac{\sum_t x_t \cdot y_t}{\sqrt{\sum_t x_t^2 \cdot \sum_t y_t^2}} \tag{6}$$

$$ED(\mathbf{x},\mathbf{y}) = \sqrt{\sum_t (x_t - y_t)^2} \tag{7}$$

$$PD(\mathbf{x},\mathbf{y}) = 1 - \frac{\sum_t (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_t (x_t - \bar{x})^2 \sum_t (y_t - \bar{y})^2}} \tag{8}$$

Here we add Dynamic Time Warping (DTW) to this range. DTW is used for measuring the dissimilarity between two time series. The technique, which is popular in speech recognition, does not require two series to have the same length. The idea of the DTW algorithm is to find the best alignment (or *warping path*) between two time series. This feature overcomes the weakness of sensitivity to distortion in the time axis compared to Euclidean distance [10].

Formally, consider two time series X and Y

$$X = x_1, x_2, \cdots, x_i, \cdots, x_m$$
$$Y = y_1, y_2, \cdots, y_i, \cdots, y_n$$

The potential alignments between X and Y can be represented in an $N \times M$ matrix where $m_{i,j}$ corresponds to the distance between x_i and y_j. The DTW distance can be computed using dynamic programming to evaluate the following recurrence:

$$d(i,j) = m_{i,j} + min\{d(i-1,j-1), d(i-1,j), d(i,j-1)\} \tag{9}$$

then

$$DTW(X,Y) = d(m,n) \tag{10}$$

C. Representing Financial News

There is a number of representations developed for the purposes of Information Retrieval that could be used in this study. These range from the simplest bag-of-words model, which only takes into account the presence (and frequency) of words in a document, but ignores any word order, to representations of words and their neighbours (bigrams, trigrams, etc.) and those in which parts of the parse tree of a sentence are used as features [11].

1) Bag-of-Words (BoW): It represents a collection of text documents as a `document` $\times$ `word` matrix which treats each word in the whole collection as a separate feature. The content of each document is then encoded as a vector containing the (relative) frequency of each of its words, including zeros for all the words that do not appear in the document. This allows for an easy comparison between any two documents, at the price of ignoring the grammatical relationship between words. So, a set of text documents $\mathcal{D}$ is represented as a matrix M where each row corresponds to a document $d \in \mathcal{D}$, and each column stands for a feature w (usually a word or token). Each element $M_{i,j}$ then is the relative frequency with which word j appears in document i.

An additional weighting scheme is often used to reduce the importance of words that appear across most documents, and highlight the ones that are characteristic to a small subset of documents. TF-IDF (Term Frequency – Inverse Document Frequency) [12] is the most popular such technique. Here the relative frequency of word w in document d is weighted according to equation 11. This reduces the perceived importance of a word w in a document d to zero if the word appears in all documents, and increases it gradually as the number of documents containing w decreases [5].

$$tfidf_{d,w} = \frac{freq_{w,d}}{|d|} \cdot log \frac{|\mathcal{D}|}{|\{d \in \mathcal{D} : w \in d\}|} \tag{11}$$

where $|\cdot|$ is size of a set; $freq_{w,d}$ is the number of occurrences of word w in document d.

2) Word2Vec: As the number of words in a large document collection could surpass 10^6 (which would result in up to 10^{12} possible bigrams, if these were used), dimension reducing techniques can also be considered in order to fight the increase in computational complexity and sparsity of data. One such approach that is quickly growing in popularity is *word2vec* [6], which uses the class of neural networks popularised under the label of Deep Learning to reduce the representation dimensionality to a value k which is typically $100 < k < 1000$. The result is that each word is represented as a linear combination of these new features, that is, a vector of size k known as *word embedding*. We then represent a document of n words as the average of its n word embeddings. A set of m documents is represented as a matrix of size $m \times k$. The method relies on the distributional statistics of words within a fixed size window. These are often collected from very large corpora and then used with other documents of interest.

3) Text Preprocessing Steps: In this study, we always preprocess all text documents in the following way. First, the text is *tokenized*, i.e. split into separate words or punctuation symbols. Then we remove all punctuation and *stop words*, essentially all pronouns, prepositions, conjunctions and a few very common verbs. The remaining words are *lemmatized*, i.e. replaced by their standard entry in the dictionary. All URLs are then mapped to the same string (URL), email addresses are mapped to the string EMAIL, and numbers are mapped to NUM. Finally, we merge all preprocessed news items for each company into a single document. From this, we produce two representations of all news on m companies making use of n different words. One is the TF-IDF weighted bag-of-words $m \times n$ matrix, the other – the $m \times k$ matrix produced with the word2vec approach (where $k = 300$).

D. Text Distance Metrics

We have previously combined the two text representations (BoW, word2vec) with two different distance metrics, CD and ED (as defined above) to produce four different distance matrices representing how the news about our companies differ from each other [3].

We now add a new distance metric, namely, the Word Mover's Distance (WMD). This is a recent document-level similarity metric [13], which defines the distance (or *travel cost*) between a pair of words from either document to be the Euclidean distance between their word embeddings. This is then used to find the closest match for each word in either document, and the normalised sum over all such pairs is then taken as the overall cost.

Here we start with the set of news items for each company, representing each news item with a single word embedding computed as the average of the embeddings of all words in that news item. WMD is then applied as described above.

Similarly, our approach first represents a set of documents as a single vector by averaging over each dimension. Therefore, a document in our approach is equivalent to a word in the WMD, and a set of documents can be treated as a sentence in the WMD metric. Since a document only occurs once in a document set, the weighting coefficient for each document is simply the inverse document frequency with respect to the size of the document set.

E. Predictive Causality Data Sampling

We use a sliding windows approach to capture the dynamic relationship between stock news and trading data. Figure 1 shows the arrangement of a sliding window of news and a sliding window for time series.

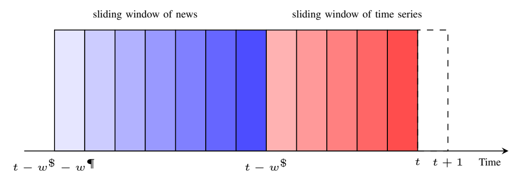

Fig. 1: A pair of sliding windows at time t: The time axis represents trading days over time. The sliding window of news in blue has $w^{¶}$ days covered and the sliding window of the time series includes $w^{\$}$ trading days. The two sliding windows do not overlap. Every subsequent pair of sliding windows advances by one trading day.

As financial markets do not trade every day, the sliding windows were generated with trading days as well as calendar days in mind. The time series window has a 5-trade-day span in order to cover a complete trading week. The news window has a 7-calendar-day span in order to cover the whole week preceding the time series window, as news is always available, no matter whether there is trade or not.

As there is no overlap between the two windows, any correlation between the news and the time series can be interpreted as a measure of predictive causality from the former to the latter.

F. Leave One Out

As suggested in the introduction, the results of either correlation test can be used to see which stock makes the greatest contribution to the result. When using consecutive, non-overlapping windows to sample the data, as described above, this can be used to identify the company for which the news has the greatest power to affect (or, at least, to be aligned with, and therefore to predict) its future financial performance.

The method we use is reminiscent of the way Sir Walter Raleigh alegedly used to *weigh smoke* in front of Queen Elizabeth I by subtracting from the weight of tobacco that of the ashes it leaves to find the weight of the missing element, smoke.

We start with the two distance matrices $\boldsymbol{M}^{¶}$ (news) and $\boldsymbol{M}^{\$}$ (time series). Our approach, which we refer to as *leave-one-out* (LOO) for obvious reasons, selects one company stock at a time and treats its contribution to the causal link between text distances and time series distances as the smoke in the above anecdote. LOO removes the i-th stock from the stock pool and recomputes the correlation between $\boldsymbol{M}^{¶}_{\text{-}i}$ and $\boldsymbol{M}^{\$}_{\text{-}i}$. The subscript '-$i$' denotes the fact that the i-th stock has been excluded. The 'weight of smoke' or change in the correlation δ_i can be obtained by subtracting $\rho_{\text{-}i}$ from ρ.

$$\delta_i = \rho - \rho_{\text{-}i} \tag{12}$$

where

$$\rho = \texttt{Correlation\_test}(\boldsymbol{M}^{\P}, \boldsymbol{M}^{\$}),$$
$$\rho_{\text{-}i} = \texttt{Correlation\_test}(\boldsymbol{M}^{\P}_{\text{-}i}, \boldsymbol{M}^{\$}_{\text{-}i})$$

`Correlation_test` can be either the Mantel test or Spearman's Rank-Order Correlation. δ_i can now be used as a technical indicator that measures the correlation contribution of target stock i to the overall portfolio.

IV. Experimental Design and Results

To measure the potential causal relationship between news and financial time series, we use the above-mentioned sampling approach and combine the components described in the previous section in the following ways:

1) Select a textual representation (Word2Vec or TF-IDF in this research) and one time series from the following alternatives: adjusted close(ADJ), change (CHG), volume (VOL), intraday return (IRT) or overnight return (ORT).
2) Pre-process news articles and time series as described in Sections II and III.C.3.
3) For each pair of time windows (see Figure 1): calculate pairwise distances between all combinations of stocks, thus obtaining two distance matrices $\boldsymbol{M}^{\P}$ and $\boldsymbol{M}^{\$}$, or 2 lists of distances for Spearman's Rank-Order Correlation.
4) For each pair of sliding windows: perform the correlation test (Mantel's or Spearman's Rank-Order Correlation) and output ρ_M, resp. ρ_S and the corresponding $p\text{-}value$.

In all cases, DTW is used to compare time series. WMD is used to compare news when represented through word embeddings, while ED is used with BoW+TF-IDF.

A. Predictive Causality Results

While we tested all relevant designs of the experiment testing for a causal link between news and time series using both correlation tests, due to space limitations we only list the results using Pearson's rank correlation (See Fig. 4). It is clear that a statistically significant causal link is present on a large proportion of days over the studied time period.

B. Evaluation through Trading

For the purpose of illustrating the effectiveness of our framework, we use a very simple trading strategy using the indicator δ described in the previous section to assist portfolio selection. After each trading day, the strategy selects the stock with the largest δ, buys the stock at next market opening and sells the stock before market closing. The daily return of the portfolio r_{t+1} can be approximated using equation 13.

$$r_{t+1} = r_{t+1}^{(n)} \tag{13}$$

where

$$n = arg\max_i\{\delta_t^{(i)}\}_{i\in\{1,\cdots,N\}}$$

n is the index of selected stock and N is the number of stocks in portfolio.

The two benchmarks are **Best Stock** (BS) and **Uniform Constant Rebalance** (UCR). BS is a strong baseline strategy which leverages market momentum. It can be briefly described as putting all capital into the stock with the best financial performance. This is a strategy solely based on time series. In this research we chose a 5-day average intraday return as the metric to measure the performance of stocks. Therefore, if stock A has the highest 5-day-average intraday return on day t after market, strategy BS will sell any valid position and buy stock A on day $t+1$ using all capital.

$$r_{t+1} = r_{t+1}^{(n)} \tag{14}$$

where

$$n = arg\max_i\{\bar{r}_t^{(i)}\}_{i\in\{1,\cdots,N\}}$$

$\bar{r}$ is the 5-day-average intraday return.

The word 'uniform' in UCR means an equal share of capital for every stock. UCR adjusts stocks and balances the capital in stocks so that they share the same portion of the total capital after every trading day [14]. The daily return of a UCR portfolio can be calculated by using the average return of all stocks in the portfolio.

$$r_{t+1} = \frac{\sum_{j=i}^{N} r_t^{(j)}}{N} \tag{15}$$

Evaluation Metrics: To determine the success of a trading strategy, the accumulated portfolio value (APV) is used for a straightforward comparison of ability to produce profit.

$$APV_t = \frac{p_t}{p_0} \tag{16}$$

where p_t is the value of assets plus funds in the portfolio at time t, p_0 is the initial fund allocated to the strategy. The higher the APV, the more profit a strategy produces.

However, APV does not give credit to those strategies which drop less when risks happen, but with a slightly lower profit. Therefore, we also calculate the Sharpe Ratio (SR) to measure the ratio between the risk and the potential profit a strategy can make.

$$SR = \frac{\mathbb{E}[r - r_F]}{\sqrt{var[r - r_F]}} \tag{17}$$

where r is the periodic return and r_F is the return of the risk-free asset. We assume the risk-free asset is cash therefore $r_F = 0$ in this research.

SR takes the volatility of the portfolio into account, but treats the upwards and downwards trends equally, even though upwards movements increase the APV of a portfolio. Maximum Drawdown (MD) is proposed to give a specific measure of how much loss at time t a portfolio can make since its maximum APV before t.

$$MD_t = \frac{max_0^{\tau}(\{APV_i\}) - min_{\tau}^{t}(\{APV_i\})}{max_0^{\tau}(\{APV_i\})} \tag{18}$$

TABLE II: Financial Performances of trading strategies. There are three groups of results: the first group shows results using the Mantel correlation; the second group shows results using Spearman's rank correlation; the last group contains two baselines: BS (the Best Stock strategy), and UCR (the Uniform-Constant Rebalance strategy). fAPV refers to final-Accumulated Portfolio Value; SR stands for Sharpe Ratio; MD is Maximum Drawdown; AR is the Annualised Return. Bold numbers labeled with '*' are the best among the participants.

News	Time series	fAPV	SR	MD [%]	AR [%]
Mantel					
W2V	ADJ	1.395	2.644	9.980	80.437
	CHG	***1.478**	2.336	15.017	***100.155**
	VOL	1.339	2.299	10.641	67.951
	IRT	1.134	0.897	11.837	24.938
	ORT	1.317	1.897	12.763	62.944
TF-IDF	ADJ	0.960	-0.273	10.911	-7.011
	CHG	1.068	0.592	17.456	12.464
	VOL	1.061	0.595	16.260	11.017
	IRT	0.929	-0.423	14.402	-12.242
	ORT	0.972	-0.121	14.805	-4.972
Spearman					
W2V	ADJ	1.164	1.471	8.233	30.870
	CHG	1.027	0.308	16.311	4.850
	VOL	1.413	***2.992**	***6.661**	84.730
	IRT	1.132	0.940	19.347	24.642
	ORT	1.209	1.483	11.461	40.146
TF-IDF	ADJ	1.151	1.368	13.937	28.268
	CHG	1.345	1.965	10.271	69.167
	VOL	1.261	1.879	15.370	50.873
	IRT	1.266	1.666	9.232	51.922
	ORT	1.339	2.244	9.857	67.930
Baseline					
	BS	1.271	2.148	12.366	53.012
	UCR	1.097	1.530	7.007	17.868

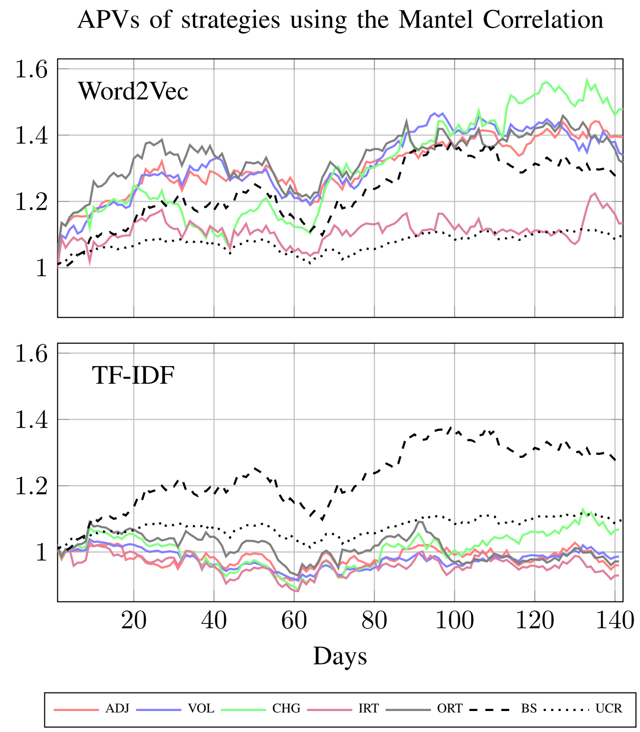

Fig. 2: Accumulated Portfolio Values (APV) of strategies using the Mantel test. The top figure shows those strategies using Word2Vec , and the bottom figure shows those using TF-IDF . Both figures contain the baseline strategies, BS (dashed line) and UCR (dotted line).

where max_0^τ means the maximum between 0 and tau and min_τ^t means the minimum between τ and t.

Annualised Return (AR) is a commonly used metric to estimate the rate of return for an investment each year.

$$AR = (\frac{p_T - p_0}{p_0})^{\frac{252}{T}} \tag{19}$$

where T refers to the number of trading days from the beginning of a strategy. 252 is a standard number of trading days in a whole calendar year.

C. Results of Portfolio Selection

Table II illustrates overall comparison of all trading strategies. Spearman Word2Vec -VOL gives the best SR (Sharpe ratio) at 2.992, however Mantel-Word2Vec combinations generally outperform the rest, especially using ADJ, CHG and VOL, whose SR varies between 2.299 and 2.644. Remarkably, the Word2Vec -VOL combinations using either statistical test yield high SR above 2.2.

Figures 2 and 3 show the APV curve over time for all strategies. The Mantel-Word2Vec combination performs better than the rest of our alternative designs and outperforms the BS strategy most of the time. Mantel-TF-IDF appear inferior to the Word2Vec representation and to the baselines. On the other hand, Spearman-based designs show diverse curves but most strategies varies between BS and UCR, except the curve of VOL.

The main goal of the evaluation experiment was to illustrate the effectiveness of the quantified relationship using the proposed framework. We compared the performance of combinations of text representations (W2V and TF-IDF), time series (ADJ, CHG, VOL, IRT, ORT) and also of correlation quantifiers (the Mantel correlation test and Spearman's rank correlation).

There were significant differences between Word2Vec and TF-IDF with regard to quantifier selection between the Mantel correlation and Spearman's Rank-Order correlation. Strategies using Word2Vec outperformed those with TF-IDF, this is in agreement with previous research by [13] that Word2Vec with WMD distance produces better clusters than TF-IDF.

The results also indicate that the choice of correlation test leads to different strategies being used. The Mantel test relies on exact distance measurements to detect linear correlation; therefore, the choice of distance metric plays an important role. Spearman's Rank Order correlation on the other hand focuses on the rankings instead of the precise value of pair-wise distances; thus, the results of Word2Vec and TF-IDF using Spearman's correlation are not significantly different. The best results of all designs are achieved for the variable *trading volume* (VOL).

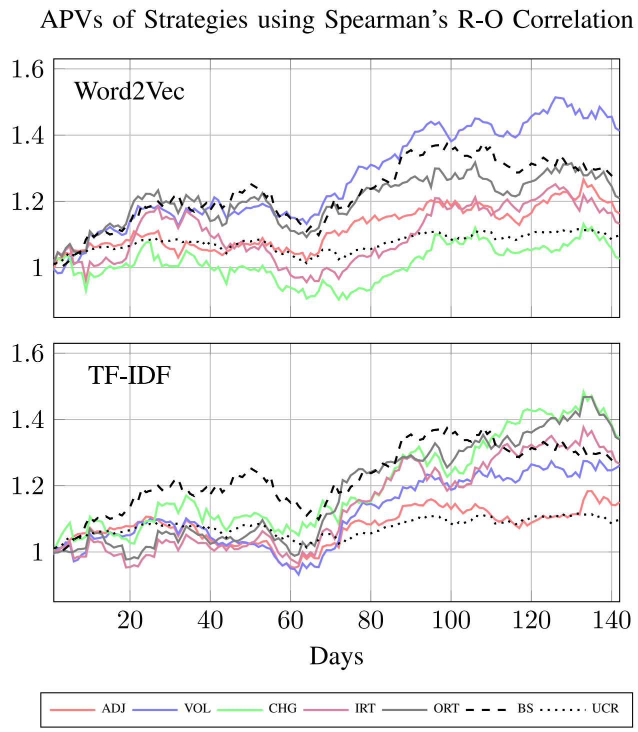

Fig. 3: Accumulated Portfolio Values (APV) of strategies using Spearman's Rank-Order correlation coefficient to quantify the causal link. The top figure shows those strategies using Word2Vec and the bottom figure shows those using TF-IDF. Both figures contain the baseline strategies, BS (dashed line) and UCR (dotted line).

V. Conclusions

The proposed framework is able to detect and quantify a causal relationship between financial news and time series without any further background knowledge.

The most remarkable result to emerge from the evaluation is that the use of LOO outperforms the strong baseline BS even when combined with a very simple trading strategy. The comparison is very revealing: despite the fact that BS bets on momentum, which is well suited to the prevailing bullish market during the period studied, it is outdone by LOO, which is not based at all on how well (or poorly) a company is performing, but only on how strongly news appears to affect that performance. We take the result as a confirmation that financial news does affect the market in a way that can inform trading decisions. Secondly, the time scale used also provides evidence that the impact of some of the news lasts longer than a day.

Generally speaking, the proposed framework has an excellent potential for practical applications. It can be combined the LOO method with ways of detecting whether the company stock it singles out is on the rise or in decline, e.g. through sentiment analysis of the relevant news, by observing the time series or through a combination of those two approaches. The result would be a fully automated trader (where incremental processing of news could make for fast reaction times), which could also be incorporated into an asset allocation algorithm for portfolio management.

In terms of future work, there is still a large number of hyper-parameters left to explore, such as the size of sliding windows for news and time series, the distance metric or the way correlation is measured.

We are aware that this research is conducted on a daily level timescale, which may prevent us from discovering shorter periods of causal links between news and time series. A further study of the effects of using hourly or even minute-by-minute data remains an interesting step to be explored.

References

[1] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *Journal of Finance*, vol. 2, pp. 383–417, May 1970.

[2] M. Butler and D. Kazakov, "Testing implications of the adaptive market hypothesis via computational intelligence," in *IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, New York, 2012.

[3] H. Qu and D. Kazakov, "Quantifying correlation between financial news and stocks," in *IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, Athens, 2016.

[4] N. Mantel, "The detection of disease clustering and a generalized regression approach," *Cancer research*, vol. 27, no. 2, pp. 209–220, 1967.

[5] J. Sedding and D. Kazakov, "Wordnet-based text document clustering," in *COLING 2004 3rd Workshop on Robust Methods in Analysis of Natural Language Data*, V. Pallotta and A. Todirascu, Eds., Geneva, 2004, pp. 104–113.

[6] T. Mikolov and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, 2013.

[7] F. Diebold, *Elements of Forecasting*, 2nd ed. Cincinnati: South Western, 2001.

[8] J. Si, A. Mukherjee, B. Liu, Q. Li, H. Li, and X. Deng, "Exploiting Topic based Twitter Sentiment for Stock Prediction," in *ACL (2)*, 2013, pp. 24–29.

[9] H. Qu, M. Sardelich, N. N. Qomariyah, and D. Kazakov, "Integrating time series with social media data in an ontology for the modelling of extreme financial events," in *LREC 2016 Joint Second Workshop on Language and Ontology & Terminology and Knowledge Structures*, Portorož, Slovenia, 2016.

[10] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and Information Systems*, vol. 7, pp. 358–386, 2005.

[11] A. Moschitti, "Efficient convolution kernels for dependency and constituent syntactic trees," in *European Conference on Machine Learning*. Springer, 2006, pp. 318–329.

[12] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information processing & management*, vol. 24, no. 5, pp. 513–523, 1988.

[13] M. J. Kusner, Y. Sun, N. I. Kolkin, and K. Q. Weinberger, "From Word Embeddings To Document Distances," *Proceedings of The 32nd International Conference on Machine Learning*, vol. 37, pp. 957–966, 2015.

[14] B. Li and S. C. H. Hoi, "Online Portfolio Selection: A Survey," 2012. [Online]. Available: http://arxiv.org/abs/1212.2129

TABLE III: Companies represented in the data set (close price shown in the graph)

Symbol	Name	Exchange	Stock Price	Number of News items
AAPL	Apple Inc.	NASDAQ		
AMZN	Amazon.com Inc.	NASDAQ		
BA	Boeing Co.	NYSE		
CMCSA	Comcast Co.	NASDAQ		
CSCO	Cisco Systems, Inc.	NASDAQ		
CVX	Chevron Co.	NYSE		
DIS	Walt Disney Co.	NYSE		
EBAY	eBay Inc.	NASDAQ		
FB	Facebook Common Stock	NASDAQ		
GE	General Electric Co.	LON		
GOOG	Alphabet Inc. Class C	NASDAQ		
GOOGL	Alphabet Inc. Class A	NASDAQ		
GS	Goldman Sachs Group Inc.	NYSE		
HD	Home Depot Inc.	NYSE		
INTC	Intel Co.	NASDAQ		
KO	The Coca-Cola Co.	NYSE		
MSFT	Microsoft Co.	NASDAQ		
NFLX	Netflix, Inc.	NASDAQ		
NKE	Nike Inc.	NYSE		
SBUX	Starbucks Co.	NASDAQ		
T	AT&T Inc.	NYSE		
TSLA	Tesla Motors Inc.	NASDAQ		
VZ	Verizon Communications Inc.	NYSE		
WMT	Wal-Mart Stores, Inc.	NYSE		
YHOO	Yahoo! Inc.	NASDAQ		

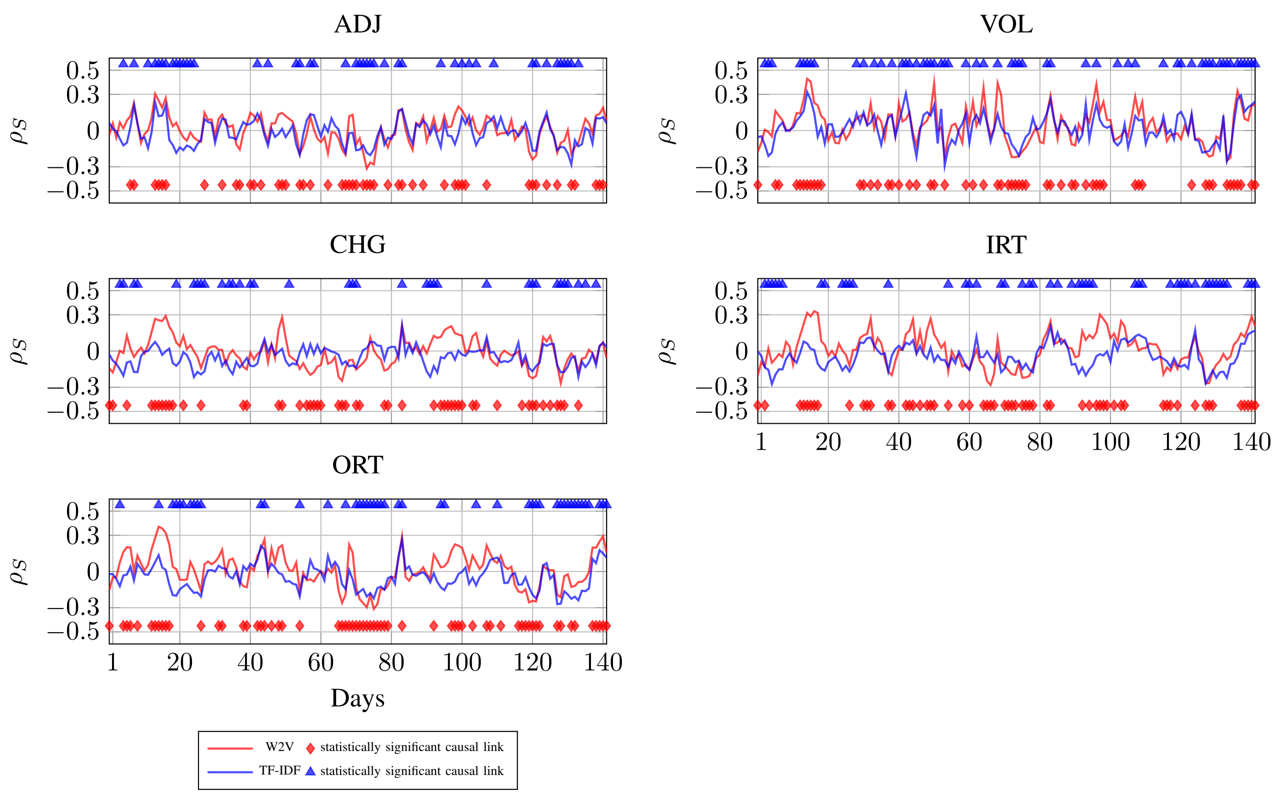

Fig. 4: Testing for a news-to-time-series causality link with Spearman's rank correlation ρ_S. The red solid line and diamonds refer to tests on distance based on W2V and the blue solid line and triangles refer to those based on BoW+TF-IDF .