International Journal of Artificial Intelligence,
ISSN 0974-0635; Int. J. Artif. Intell.
2019, Volume 17, Issue Number 2

Predicting a Stock Portfolio with the Multivariate Bayesian Structural Time Series Model: Do News or Emotions Matter?

S. Rao Jammalamadaka[1], **Jinwen Qiu**[1], **Ning Ning**[2]

[1]Dept. of Statistics & Applied Probability
University of California, Santa Barbara
rao@pstat.ucsb.edu, jqiu@pstat.ucsb.edu

[2]Dept. of Applied Mathematics
University of Washington, Seattle
ningnin@uw.edu

ABSTRACT

In this paper, we provide methods for creatively incorporating information from financial news and Twitter feeds into predicting the prices of a portfolio of stocks, using the framework of the Multivariate Bayesian Structural Time Series (MBSTS) model. MBSTS is a Bayesian machine learning model designed to capture correlations among multiple target time series, while using a number of contemporaneous predictors. As an illustration of the current model, we use data on two leading online commerce companies, namely Amazon and eBay, and run extensive empirical experiments to examine which if any, text mining predictors would add to the predictability of a stock price. Evaluation of competing models such as the autoregressive integrated moving average (ARIMA) model, and the recurrent neural network (RNN) model with long short term memory (LSTM), in terms of their performances with respect to cumulative one-step-ahead forecast errors with and without sentimental predictors, were carried out. Our contributions are threefold: Firstly, our model is the first one that successfully incorporated the online text mining to an advanced multivariate Bayesian machine learning time series model, which opens the door of applying both text mining and machine learning simultaneously in modern quantitative finance research; Secondly, under the presence of both modern and classical predictors in both fundamental and technical sense, the polarity of news still adds on a complementary effect; Thirdly, we discover that all models under investigation with sentimental predictors do outperform models without these sentimental predictors, and the MBSTS model with sentimental predictors outperforms all the other models.

Keywords: Feature Selection, Time Series Forecast, Text Mining, Sentiment Analysis.

2010 Mathematics Subject Classification: 68T01, 37M10, 62F15.

2012 ACM Computing Classification System: Computing methodologies–Artificial intelligence–Natural language processing–Lexical semantics;500, Computing methodologies–Machine learning–Machine learning algorithms–Feature selection;500.

1 Introduction

1.1 Importance of Sentiment Analysis

Text data mining, also known as text mining, refers to the process of extracting non-trivial patterns and high-quality information from unstructured text documents, and is an extension of knowledge discovery from unstructured databases. It usually involves the process of structuring the input text, deriving patterns within the structured data, and finally evaluating and interpreting the output, with the overarching goal of turning the text into data for analysis via applications of natural language processing and analytical methods. Specifically, text analysis involves information retrieval, lexical analysis, pattern recognition, annotation, information extraction, data mining, visualization, and predictive analytics. Emotion artificial intelligence, also known as sentiment analysis or opinion mining, as a typical text mining task, refers to the use of natural language processing, text analysis, and computational linguistics to systematically identify, extract, quantify and study effective states and subjective information. It aims to determine the attitude of a speaker or writer with respect to the overall contextual polarity or emotional reaction to a document.

Nowadays, unstructured text data on the web and in social media communications is rapidly becoming a great source of useful information, ranging from news articles to personal opinions such as Twitter feeds. Over the past several years, significant progress has been made in sentiment classification, opinion recognition, and opinion analysis, including the following: A personalized news reader enhanced by machine learning and semantic filtering was created in (Banos, Katakis, Bassiliades, Tsoumakas and Vlahavas, 2006); Specifications of knowledge components for reuse was investigated in (Motta, Fensel, Gaspari and Benjamins, 1999); Ontology-based sentiment analysis of Twitter posts was performed in (Kontopoulos, Berberidis, Dergiades and Bassiliades, 2013); Combining textual and semantic descriptions for automated semantic web service classification was executed in (Katakis, Meditskos, Tsoumakas and Bassiliades, 2009); An infrastructure to support cooperation of knowledge-level agents on the semantic Grid was proposed in (Dragoni, Gaspari and Guidi, 2006); A distributed focused crawler to support open research with Twitter data named Twitterecho was invented in (Bošnjak, Oliveira, Martins, Mendes Rodrigues and Sarmento, 2012); A framework and infrastructure for semantic web services was proposed in (Motta, Domingue, Cabral and Gaspari, 2003); Automatically creating a reference corpus for political opinion mining in user-generated content was explored in (Sarmento, Carvalho, Silva and De Oliveira, 2009); Tokenizing micro-blogging messages using a text classification approach was proposed in (Laboreiro, Sarmento, Teixeira and Oliveira, 2010); Searching for clues to detect irony in user-generated contents was performed in (Carvalho, Sarmento, Silva and De Oliveira, 2009).

1.2 Multivariate Bayesian Structural Time Series (MBSTS) Model

Nowadays, machine learning, as a part of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to learn from data, is changing virtually every aspect of our lives. However, typical machine learning assumptions such as

data being independent and identically distributed, are not satisfactory when dealing with time-series data with multiple predictors. In (Scott and Varian, 2014) and (Scott and Varian, 2015), a Bayesian structural time series (BSTS) model was introduced, by adapting a new Bayesian machine learning technique to a given time series along with possible covariates. The BSTS model has recently been extended by (Qiu, Jammalamadaka and Ning, 2018) who proposed a multivariate Bayesian structural time series (MBSTS) model for dealing with multiple target time series, which helps in feature selection and forecasting in the presence of related external information. The Bayesian paradigm in the multivariate setting avoids overfitting and is able to take advantage of the correlations among multiple target time series. Remarkably, the MBSTS model is able to provide the most desired flexibility in selecting a different set of predictors for each target series.

The MBSTS model has three main features:

- **Hierarchical models** are a type of linear regression models in which the observations fall into hierarchical or completely nested levels. Different techniques have been adopted for the purpose of multilevel modeling (see (Tantrum, Murua and Stuetzle, 2004) and the references therein), and have been evaluated and assessed (see (Tantrum, Murua and Stuetzle, 2003)). The MBSTS model builds the hierarchical structure by stochastically decomposing a time series into suitable components such as trend, seasonality, and regression on covariates, with the help of Kalman filter (see (Harvey, 1990), (Durbin and Koopman, 2002), and (Petris, Petrone and Campagnoli, 2009)).

- **Machine learning** is incorporated in the regression component of the MBSTS modeling structure. Modern regression analysis using machine learning technique has been well developed over the past decade (see (Slini, Karatzas and Papadopoulos, 2002), (Pai and Lin, 2005), (Pai, Lin, Lin and Chang, 2010), and (Lin, Pai, Lu and Chang, 2013)). Different to other machine learning techniques adopted in some recent time-series analysis works (for example, (Glezakos, Tsiligiridis, Iliadis, Yialouris, Maris and Ferentinos, 2009), (Koutroumanidis, Iliadis and Sylaios, 2006), (Voukantsis, Karatzas, Kukkonen, Räsänen, Karppinen and Kolehmainen, 2011), (Yan, Qiu and Xue, 2009), (Yan, 2012), and (Wang, Yan and Oates, 2017)), the MBSTS model uses the “spike and slab” variable selection technique (see (George and McCulloch, 1997) and (Madigan and Raftery, 1994)) to fulfill the mission of feature selection in the regression component of multivariate time series.

- **Forecasting** is one of the main purposes of time series analysis, but it is a dark horse in the field of data science (see (Slini, Karatzas and Moussiopoulos, 2002), (Athanasiadis, Karatzas and Mitkas, 2006), (Kasabov and Song, 2002), and (Pai, Wei-Chiang, Ping-Teng and Chen-Tung, 2006) and the references therein in for description and applications of different corresponding techniques). The MBSTS model uses the Bayesian model averaging technique (see (Hoeting, Madigan, Raftery and Volinsky, 1999)) to combine all the feature selection results in predicting future values, whose superior forecasting performances over benchmark time series models were verified through extensive simulations and real data experiments.

1.3 Embed Sentiment Analysis in the MBSTS Model

Motivation: Over the past several years, significant progress has been made in sentiment classification, opinion recognition, and opinion analysis, which extracts indicators and/or provide representations of public moods directly from social media contents such as blog contents. Although it has been widely used in judging customer reviews and survey responses online and in social media, in contexts such as marketing and customer service, its applications in the financial industry are still at a developing stage, let alone be embed into a financially applicable machine learning framework. Widely acknowledged, news articles about a specific company can spread information and influence people either consciously or unconsciously in their decision making. Good news about one company usually pushes up its stock price, while bad news has the effect of reducing shareholders' confidence to sell their holdings and then leads to a decreased stock price. In this paper, we use retrieved sentimental information from financial news and Twitter feeds in multivariate stock price time series prediction, in the framework of the MBSTS model.

Approach: We adopt a lexicon-based approach for sentiment classification in terms of polarities (*positive, negative, neutral*) of news articles, and emotions (*anxiety, calmness, dislike, fear, liking, love, joy, sadness, unknown*) of Twitter posts. The lexicon-based approaches for sentiment classification have the advantage of avoiding the cumbersome step of labeling training data, and are based on the insight that sentiments conveyed by a piece of text can be obtained on the basis of the polarities and emotions of the words which comprise it. We conduct an empirical study with one-step-ahead predictions on the max log returns of two leading e-commerce companies, namely Amazon and eBay. For each company, besides 27 Google domestic trends and 8 stock technical predictors, we bring in 2 polarity and 8 emotion predictors, from which to perform feature selection in the regression component of the MBSTS modeling structure.

Main findings: In terms of the posterior inclusion probabilities, our feature selection analysis found that even in the presence of Google domestic trends and the standard stock price technical predictors (see (Qiu et al., 2018) for discussion on the sufficiency and necessity in using these two types of predictors for financial time series forecasting), polarity predictors should still be included according to the model training results, while emotion predictors did not have the same significant impacts on stock prices. We further verified that the polarity predictors alone played a crucial role in explaining the fluctuations that were not otherwise explained by the local trend component. Model performances with and without sentimental predictors in terms of cumulative one-step-ahead forecast errors were examined, and the results revealed that sentimental predictors consistently increased the forecasting power. One-step-ahead predictions with 80% confidence intervals over a month indicate that the predictions closely match the changing patterns of the target time series.

Model performance: We compared the model performance with the autoregressive integrated moving average (ARIMA) model which is a classical and benchmark time series model, and the recurrent neural network (RNN) model with long short term memory (LSTM) which is a successful artificial neural network model well-suited to classifying, processing, and making predictions based on time series data. We fully examined the case with and without sentimen-

tal predictors, as well as the basic models without any predictors. Our exhaustive analyses lead to the following conclusions: First, all models with sentimental predictors outperform models without sentimental predictors, and further outperform the standard models without any predictors; Second, the MBSTS model with sentimental predictors outperforms all the other models.

1.4 Organization of the Paper

The rest of the paper is organized as follows. In Section 2, we perform sentiment analysis in terms of polarities and emotions. In Section 3, we provide the methodology on how to properly embed sentiment analysis in the MBSTS model. In Section 4, we investigate whether news and public emotions have any significant effect on the stock prices in the framework of the MBSTS model; further comparisons to other benchmark models are done showing that the MBSTS model with sentiment indicators included, outperforms them all. In Section 5, we make some concluding comments.

2 Sentiment Analysis

Sentiment analysis involves discerning subjective materials and extracting various forms of attitudinal information. In this section, we perform sentiment analysis in terms of polarities and emotions.

2.1 Polarities

Since there is no natural labeling of the datasets, we apply an unsupervised approach to determine the polarity of each textual document. A popular unsupervised technique is to utilize discriminatory-word lexicons, which uses dictionaries of words annotated with the word's semantic orientation or polarity. Lexicon-based approaches for sentiment classifications are based on the insight that the polarity of a textual document can be determined by the polarities of all words which compose it. The lexicon resource used in this paper is SenticNet, which provides a set of semantics, sentics, and polarities associated with $100,000$ natural language concepts. As a knowledge base, SenticNet is able to identify polarity and affective information, including opinion mining on complex concepts such as accomplishing goals, celebrating special occasions, losing temper and so on. Polarity scores in the range of -1 to 1 for these common sense concepts are provided by this lexicon.

In this literature, inspired by (Musto, Semeraro and Polignano, 2014) and the references therein, a more fine-grained approach containing five steps to take into account many facets of language such as negation, is implemented. At the first step, we split a specific company's one news article D_i into sentences s_{ij}, according to the splitting cues of punctuations such as periods and question marks. At the second step, we filter out the irrelevant sentences while keeping only those containing items such as stock symbols, company names, and product names. At the third step, a polarity score for each sentence is calculated by a weighted sum

of the polarity scores of phrases or words $\{t_{ijk}\}$. Whenever a negation is identified in the sentence, we invert the polarity of this sentence by changing the sign of its original score. In line with (Musto et al., 2014), the polarity of the sentence s_{ij} is calculated as

$$pol(s_{ij}) = \sum_{k=1}^{m} \frac{score(t_{ijk}) * w_{pos(t_{ijk})}}{|s_{ij}|}, \tag{2.1}$$

where $score(t_{ijk})$ is the polarity score of word t_{ijk} assigned by SenticNet, $|s_{ij}|$ is the number of words in the sentence s_{ij}, and a bigger weight is assigned to the word t_{ijk} belonging to specific part of speech (POS) categories as follows

$$w_{pos(t_{ijk})} = \begin{cases} 1.5, & pos(t_{ijk}) \in \{\text{adverbs, verbs, adjectives}\}, \\ 1, & \text{otherwise.} \end{cases} \tag{2.2}$$

At the fourth step, the polarity score of each news article can be computed as the sum of polarity scores of all sentences in that news article, i.e.,

$$pol(D_i) = \sum_{j=1}^{n} pol(s_{ij}). \tag{2.3}$$

Finally, based on that news article's sentiment score, we classify it into three categories: positive, negative, and neutral. Specifically, if the sentiment score is higher than a user-defined positive threshold, then it will be classified as *positive*; if the sentiment score is lower than a user-defined negative threshold, then it will be treated as *negative*; otherwise, it will be considered as *neutral*. The positive (resp. negative) threshold can be a summary statistic, such as the median of all positive (resp. negative) polarity scores.

2.2 Emotions

Most works on sentiment analysis focus on the polarity classification while ignoring the rich and multi-dimensional structure of human emotions. Behavioral economics tells us that emotions can profoundly affect individual behavior and decision making, and recent studies reveal that changes in public mood carried in Twitter feeds are able to effectively predict stock market movements several days in advance (see, the case study of the Dow Jones Industrial Average values in (Bollen, Mao and Zeng, 2011)). To capture the dimension effects associated with public emotions, instead of using the standard 6 dimensions, we classify the Twitter feeds into 9 categories: *anxiety, calmness, dislike, fear, liking, love, joy, sadness*, and *unknown*. The classification is implemented through the WordNet-Affect lexicon proposed in (Strapparava and Valitutti, 2004), an affective extension of WordNet, including a subset of synsets suitable to represent affective concepts correlated with affective words. The WordNet-Affect lexicon extends the original linguistic resources with a set of additional affective labels (A-Labels) and a deeper hierarchical organization, in order to specialize synsets with more emotion indications, such that each item is not only labeled by its orientation (positive or negative) but also by its emotions' category.

Similar to the polarity calculation, we firstly separate Twitter feeds $\{T_i\}$ into micro-phrases $\{m_{ij}\}$ according to the standard splitting cues (punctuations, conjunctions), and then further

map each term t_{ijk} in the micro-phrase m_{ij} to its respective emotion category in the WordNet-Affect lexicon. Eight emotion scores of each micro-phrase are calculated by the weighted sum of indicator functions of mapped categories in the lexicon, as follows:

$$E_v(m_{ij}) = \sum_{k=1}^{m} \frac{E_v(t_{ijk}) * w_{pos(t_{ijk})}}{|m_{ij}|}, \quad \text{for } v \in \{\textit{anxiety, calmness, dislike, fear, liking, love, joy, sadness}\}, \tag{2.4}$$

where $E_v(t_{ijk})$ is defined, to tell whether the mapped emotion category $M(t_{ijk})$ is v, as

$$E_v(t_{ijk}) = \begin{cases} 1, & M(t_{ijk}) = v, \\ 0, & \text{otherwise}, \end{cases} \tag{2.5}$$

and $w_{pos(t_{ijk})}$ is defined in equation (2.2). Furthermore, when negation is detected in the micro-phrase, its mapped category will be changed to the corresponding antonym category (such as, *liking* $\rightarrow$ *dislike*, *sadness* $\rightarrow$ *joy*). Finally, eight emotion scores of each Tweeter feed T_i are calculated by summing up the corresponding emotion scores of its micro-phrases:

$$E_v(T_i) = \sum_{j=1}^{n} E_v(m_{ij}). \tag{2.6}$$

The classification criterion is to assign an emotion label to a specific Twitter feed, based on its maximum emotions score, i.e. the label of T_i is assigned to the optimizer

$$v^* := \arg\max_v E_v(T_i) \text{ for } v \in \{\textit{anxiety, calmness, dislike, fear, liking, love, joy, sadness}\}. \tag{2.7}$$

In the case that eight emotion scores of a specific tweet feed are all zero, its label will be assigned to *unknown*.

3 Methodology

In this section, we provide a methodology on embedding sentiment analysis in the MBSTS model framework. In order to better demonstrate the methodology and further investigate the irreplaceable effects that sentiment analysis on financial time series prediction, even in the presence of other so-called fully sufficient and necessary predictors, we illustrate the approach through empirical data analysis of two competing electronic commerce giants Amazon and eBay whose stock prices have strong correlations. Clearly, these two world-famous companies have always attracted a great deal of people's attention in the form of financial news and Twitter feeds. Daily data of stock prices were obtained from Yahoo! Finance; Financial news articles were obtained from both Google Finance and Yahoo! Finance; Twitter feeds were obtained from Twitter. We split the dataset into a training set $(01/04/2016 - 01/16/2018)$ and a test set $(01/17/2018 - 02/13/2018)$.

In Section 3.1, we build up the MBSTS model framework by stochastically decomposing the 2-dimensional target time series of Amazon and eBay, into a linear trend component and a regression component. In Section 3.2, we generate the fundamental time series predictors from Google Domestic Trends, technical time series predictors from well-recognized industry

technical indicators, and sentimental time series predictors generated by the lexicon method in Section 2. In Section 3.3, we provide the model training Algorithm 1 incorporating model setup and feature selection among all candidate time series predictors in the Bayesian paradigm, and prediction Algorithm 2 by means of Bayesian modeling averaging.

3.1 Model Framework Setup

The MBSTS model uses machine learning techniques for feature selection, time series forecasting, nowcasting, and other applications, and fully takes into consideration of the correlations among different target series. While the MBSTS model can decompose the m-dimensional target time series $\tilde{y}_t$ in several different components, in this study we only consider a linear trend component $\tilde{\mu}_t$ and a regression component $\tilde{\xi}_t$, for the reason that this is already sufficiently satisfactory for our purposes

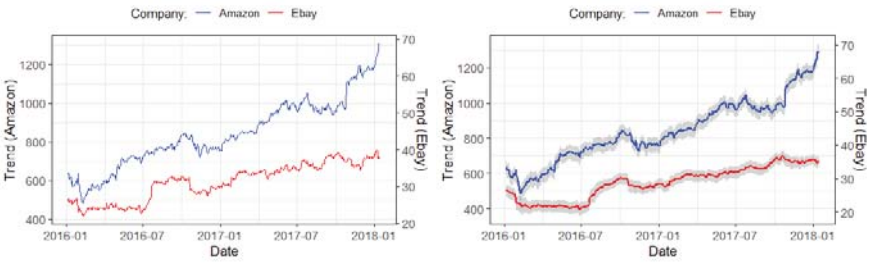
$$\tilde{y}_t = \tilde{\mu}_t + \tilde{\xi}_t + \tilde{\epsilon}_t, \qquad \tilde{\epsilon}_t \overset{iid}{\sim} N_2(0, \Sigma_\epsilon), \quad t = 1, 2 \ldots, n, \tag{3.1}$$

where $\tilde{\epsilon}_t$ represents the observation error terms.
We consider the local linear trend component in the form as:

$$\tilde{\mu}_{t+1} = \tilde{\mu}_t + \tilde{\delta}_t + \tilde{u}_t, \qquad \tilde{u}_t \overset{iid}{\sim} N_2(0, \Sigma_\mu), \qquad \Sigma_\mu \sim IW(w_\mu, W_\mu), \tag{3.2}$$

$$\tilde{\delta}_{t+1} = \tilde{\rho}\tilde{\delta}_t + \tilde{v}_t, \qquad \tilde{v}_t \overset{iid}{\sim} N_2(0, \Sigma_\delta), \qquad \Sigma_\delta \sim IW(w_\delta, W_\delta). \tag{3.3}$$

Here, $\tilde{\delta}_t$ is the expected increase in $\tilde{\mu}_t$. The parameter $\tilde{\rho}$ is a 2×2 diagonal matrix, whose diagonal entries take values in $[0, 1]$ representing the learning rates at which the local trend is updated for each target series; in this study, we set the diagonal entries to 0.7 and 1 according to the results of hyper-parameter tuning. Figure 1(b) shows the posterior distributions of the local linear trend, which capture general upward patterns of stock prices in Figure 1(a). The detrend series defined by the differences between stock prices and the mean of trend distributions are left to be explained by a pool of candidate predictors.



(a) Stock Price (b) Trend

Figure 1: Stock price and a local linear trend with 80% confidence band

The regression component with static coefficients is written as follows:

$$\tilde{\xi}_t = \{\xi_t^{(i)}\}^T, \quad \xi_t^{(i)} = \beta_i^T x_t^{(i)}, \quad i = 1, 2. \tag{3.4}$$

Here, for each target series $y^{(i)}$, $x_t^{(i)} = [x_{t1}^{(i)}, \ldots, x_{tk_i}^{(i)}]^T$ is the pool of all available predictors at time t, and $\beta_i = [\beta_{i1}, \ldots, \beta_{ij}, \ldots, \beta_{ik_i}]^T$ represents corresponding static regression coefficients. We expect a high degree of sparsity in feature selection in the sense that coefficients of the vast majority of predictors would be zero. A natural way to represent such sparsity in the Bayesian paradigm is through the "spike and slab" coefficients, hence we define $\gamma = [\gamma_1, \gamma_2]$ where $\gamma_i = [\gamma_{i1}, \ldots, \gamma_{ik_i}]$, and then set $\gamma_{ij} = 1$ if $\beta_{ij} \neq 0$ and $\gamma_{ij} = 0$ if $\beta_{ij} = 0$. Denote β_γ as the subset of elements of β where $\beta_{ij} \neq 0$, and let X_γ be the subset of columns of X where $\gamma_{ij} = 1$. The "spike" prior is written as:

$$\gamma \sim \prod_{i=1}^{2} \prod_{j=1}^{k_i} \pi_{ij}^{\gamma_{ij}} (1 - \pi_{ij})^{1-\gamma_{ij}}, \tag{3.5}$$

where π_{ij} is the prior inclusion probability of the j-th predictor for the i-th target time series which would be simplified to set a single value for all j as $\pi_{ij} = \pi_i$. In this study, we assigned prior inclusion probabilities 0.16 to Amazon and 0.10 to eBay and this prior setup would be confirmed by posterior results given later.

A simple "slab" prior specification is to make the priors on β and Σ_ϵ conditionally independent (see (Griffiths, 2003)):

$$\begin{aligned} p(\beta, \Sigma_\epsilon, \gamma) &= p(\beta|\gamma)p(\Sigma_\epsilon|\gamma)p(\gamma), \\ \beta|\gamma \sim N_K(b_\gamma, A_\gamma^{-1}), &\quad \Sigma_\epsilon|\gamma \sim IW(v_0, V_0), \end{aligned} \tag{3.6}$$

where b_γ is the vector of prior means and $A_\gamma = \kappa(\omega X_\gamma^T X_\gamma + (1 - \omega) diag(X_\gamma^T X_\gamma))/n$ is the full-model prior information matrix, with κ the number of observations worth of weight on the prior mean vector b_γ. Here, $IW(v_0, V_0)$ is the inverse Wishart distribution with v_0 the number of degrees of freedom and V_0 a 2×2 scale matrix, where $V_0 = (v_0 - 3) * (1 - R^2) * \Sigma_y$ and Σ_y is the variance-covariance matrix for multiple target time series Y. In this study, according to the results of hyper-parameter tuning, we set $\kappa = 1$, $w = 0.01$, $b = 0$, $v_0 = 4$ and $R^2 = 0.8$.

3.2 Predictors Setup

In the following, we introduce three kinds of predictors: fundamental predictors, technical predictors, and sentimental predictors. Fundamental predictors are the so-called "Google Domestic Trends" which have been developed and successfully tested in recent years; technical predictors are well recognized in Wall Street with predetermined formulas, and can be treated as classical predictors; while sentimental predictors are what we are currently investigating, whose strengths and validity will be checked after the inclusion of the above two kinds of predictors.

3.2.1 Fundamental Predictors

Fundamental analysis is a method of evaluation by attempting to measure the intrinsic value of a stock, which incorporates everything from the overall economic status to industry conditions to the financial management of specific companies. That is, fundamental analysis includes

Trend	Abbr.	Trend	Abbr.
Advertising & marketing	advert	Air travel	airtvl
Auto buyers	auto	Auto financing	autoby
Automotive	autofi	Business & industrial	bizind
Bankruptcy	bnkrpt	Commercial Lending	comlnd
Computers & electronics	comput	Construction	constr
Credit cards	crcard	Durable goods	durble
Education	educat	Finance & investing	invest
Financial planning	finpln	Furniture	furntr
Insurance	insur	Jobs	jobs
Luxury goods	luxury	Mobile & wireless	mobile
Mortgage	mtge	Real estate	rlest
Rental	rental	Shopping	shop
Small business	smallbiz	Travel	travel
Unemployment	unempl		

Table 1: Google Domestic Trends

economic analysis, industry analysis, and company analysis. For industry and company analysis, financial statements for companies are usually released to public quarterly or annually. Apparently, it is impossible to obtain crucial information on a daily basis for any of these fundamental analyses. Since 2004, Google developed the database "Google Domestic Trends" to collect the daily volume of searches related to various aspects of economics. The correlations between Google domestic trends and the equity markets have been acknowledged, and Google domestic trends have been used as representations of various economic factors (see (Preis, Moat and Stanley, 2013) and (Qiu et al., 2018)). In this empirical study, we used 27 Google domestic trends shown in the following Table 1 along with their abbreviations.

3.2.2 Technical Predictors

Technical analysis is a method that recognizes the patterns and trends in the historical prices and volumes to forecast accordingly, which is under the assumption that useful information is already reflected in stock prices. We selected a representative set of technical indicators to capture volatilities, close location values, potential reversals, momentums, and trends of stocks. In this empirical study, the formulas of 8 technical predictors in Table 2, were applied to the company data, and then 16 different technical predictors for Amazon and eBay were generated.

3.2.3 Sentimental Predictors

Evidence supports the hypothesis that sentiments reflected in mass communications and social media can be helpful in predicting stock price fluctuations, whether up or down (see, (Bollen et al., 2011)). To investigate and predict daily variations of stock prices, we collected 10 most

Variable	Abbr.
Chaikin volatility	ChaVol
Yang and Zhang Volatility historical estimator	Vol
Arms' Ease of Movement Value	EMV
Moving Average Convergence/Divergence	MACD
Money Flow Index	MFI
Aroon Indicator	AROON
Parabolic Stop-and-Reverse	SAR
Close Location Value	CLV

Table 2: Stock Technical Predictors

related financial news for each company everyday by using Google search API, retrieved 500 Twitter feeds for each company using keywords search queries such as "Amzn", "Amzon", and "eBay" also on a daily basis, and further classified them into several categories by means of the sentiment analysis method in Section 2. Two predictors representing market directional information were created by normalized values of the sum of positive or negative scores from financial news for Amazon and eBay respectively, as shown in Table 3. Eight public emotion variables, namely *anxiety, calmness, dislike, fear, liking, love, joy* and *sadness*, are measured from Twitter feeds to represent general public emotions, as shown in the Table 3.

Amazon	Abbr.	eBay	Abbr.
Positive scores for Amazon	pos.amzn	Positive scores for eBay	pos.eBay
Negative scores for Amazon	neg.amzn	Negative scores for eBay	neg.eBay
Anxiety socres for Amazon	anx.amzn	Anxiety socres for eBay	anx.eBay
Calmness scores for Amazon	cal.amzn	Calmness scores for eBay	cal.eBay
Dislike scores for Amazon	dis.amzn	Dislike scores for eBay	dis.eBay
Fear scores for Amazon	fear.amzn	Fear scores for eBay	fear.eBay
Liking scores for Amazon	lik.amzn	Liking scores for eBay	lik.eBay
Love scores for Amazon	love.amzn	Love scores for eBay	love.eBay
Joy scores for Amazon	joy.amzn	Joy scores for eBay	joy.eBay
Sadness scores for Amazon	sad.amzn	Sadness scores for eBay	sad.eBay

Table 3: Sentiment Predictors

3.3 Theory and Algorithms

For $\tilde{y}_t$, $\tilde{\mu}_t$, and $\tilde{\epsilon}_t$ in equation (3.1), set $Y = [\tilde{y}_1, \cdots, \tilde{y}_n]^T$, $M = [\tilde{\mu}_1, \cdots, \tilde{\mu}_n]^T$, and $E = [\tilde{\epsilon}_1, \cdots, \tilde{\epsilon}_n]^T$. Then we can rewrite the model in a long matrix form

$$\tilde{Y} = \tilde{M} + X\beta + \tilde{E}, \tag{3.7}$$

where $\tilde{Y} = vec(Y)$, $\tilde{M} = vec(M)$, $\tilde{E} = vec(E)$, $\beta = [\beta_1, \beta_2]^T$, and $X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}$ with matrix X_i representing all observations of the candidate predictors for the time series i. We denote

$$\tilde{Y}^\star = \tilde{Y} - \tilde{M} \tag{3.8}$$

as the multiple target time series $\tilde{Y}$ with trend time series component subtracted out. We further denote

$$\hat{X} = ((U^{-1})^T \otimes I_n)X \tag{3.9}$$

and

$$\hat{Y}^\star = ((U^{-1})^T \otimes I_n)\tilde{Y}^\star \tag{3.10}$$

where U satisfies $\Sigma_\epsilon = U^T U$.

The MBSTS model uses the Markov chain Monte Carlo (MCMC) method, which is to sample from a probability distribution based on constructing a Markov chain that has the desired distribution as its equilibrium distribution, to perform model stochastic decomposition and feature selection. Through looping through the five steps given in Algorithm 1, a sequence of draws $\tilde{\psi} = (\tilde{\mu}, \tilde{\delta}, \Sigma_\mu, \Sigma_\delta, \gamma, \Sigma_\epsilon, \beta)$ can be generated from the posterior distribution of the model, which forms a Markov chain with stationary distribution.

After modification to fit the current setting, we provide the results in (Qiu et al., 2018), where the subscript γ is used to denote the corresponding result after feature selection such as $\hat{X}_\gamma$, as follows:

- The conditional distribution of the variance-covariance matrix Σ_u in the error term of the trend component in equations (3.2) and (3.3), for $u \in \{\mu, \delta\}$, given $\tilde{Y}$ in equation (3.7) and u, is given by

$$\Sigma_u | \tilde{Y}, u \sim IW(w_u + n, W_u + AA^T), \tag{3.11}$$

  with matrix A representing a collection of residues of each time series component.

- The conditional distribution of the indicator variable γ in equation (3.5), given the variance-covariance matrix Σ_ϵ of the observation error term in equation (3.1) and $\tilde{Y}^\star$ in equation (3.8), can be expressed as

$$p(\gamma | \Sigma_\epsilon, \tilde{Y}^\star) \propto \frac{|A_\gamma|^{1/2} p(\gamma)}{|\hat{X}_\gamma^T \hat{X}_\gamma + A_\gamma|^{1/2}} \exp\left(-\frac{1}{2}\{b_\gamma^T A_\gamma b_\gamma - Z_\gamma^T (\hat{X}_\gamma^T \hat{X}_\gamma + A_\gamma)^{-1} Z_\gamma\}\right), \tag{3.12}$$

  where

$$Z_\gamma = (\hat{X}_\gamma^T \hat{Y}^\star + A_\gamma b_\gamma). \tag{3.13}$$

- The conditional distribution of the regression coefficient β in equation (3.4), given $\hat{Y}^\star$ in equation (3.10), Σ_ϵ, and γ, is given by

$$\beta | \hat{Y}^\star, \Sigma_\epsilon, \gamma \sim N(\tilde{\beta}_\gamma, (\hat{X}_\gamma^T \hat{X}_\gamma + A_\gamma)^{-1}), \tag{3.14}$$

  where

$$\tilde{\beta}_\gamma = (\hat{X}_\gamma^T \hat{X}_\gamma + A_\gamma)^{-1} (\hat{X}_\gamma^T \hat{Y}^\star + A_\gamma b_\gamma). \tag{3.15}$$

- The conditional distribution of Σ_ϵ, given $\tilde{Y}^\star$, β, and γ, is given by

$$\Sigma_\epsilon|\tilde{Y}^\star, \beta, \gamma \sim IW(v_0 + n, E_\gamma^T E_\gamma + V_0). \tag{3.16}$$

  where v_0 and V_0 are the prior distribution setup parameters in equation (3.6), n is the number of observations in each time series, and

$$E_\gamma = Y - M - X_\gamma B_\gamma. \tag{3.17}$$

Algorithm 1 incorporates model stochastic decomposition and feature selection. (Qiu et al., 2018) only provided a theoretical foundation on how to perform model forecast. In Algorithm 2, we give a precise algorithm on that. Although the core of both forecast methods are based on the posterior predictive distribution and is consistent in the Bayesian paradigm used in model training, our method is naturally able to forecast several days ahead instead of just one.

Algorithm 1 Model Training

1: Draw $(\tilde{\mu}, \tilde{\delta})$ based on the model setup, given model parameters and $\tilde{Y}$, namely $p(\tilde{\mu}, \tilde{\delta}|\tilde{Y}, \Sigma_\mu, \Sigma_\delta, \gamma, \Sigma_\epsilon, \beta)$, using the posterior simulation algorithm given in (Durbin and Koopman, 2002).
2: Draw Σ_u for $u \in \{\mu, \delta\}$ according to the conditional distribution $\Sigma_u \sim p(\Sigma_u|\tilde{Y}, u)$ in equation (3.11).
3: Draw γ_i given $(\gamma_{-i}, \tilde{Y}^\star, \Sigma_\epsilon)$ for a random i looping through the indexes of all the candidate predictors (fundamental predictors, technical predictors, and sentimental predictors in Section 3.2), based on the conditional distribution $\gamma \sim p(\gamma|\tilde{Y}^\star, \Sigma_\epsilon)$ in equation (3.12), using the stochastic search variable selection (SSVS) algorithm given in (George and McCulloch, 1997).
4: Draw β according to the conditional distribution $\beta \sim p(\beta|\Sigma_\epsilon, \gamma, \hat{Y}^\star)$ in equation (3.14).
5: Draw Σ_ϵ according to the conditional distribution $\Sigma_\epsilon \sim p(\Sigma_\epsilon|\gamma, \tilde{Y}^\star, \beta)$ in equation (3.16).

Algorithm 2 Model Forecast

1: Draw the next latent time series states $\alpha_{t+1} = (\tilde{\mu}_{t+1}, \tilde{\delta}_{t+1})$, given current states $\alpha_t = (\tilde{\mu}_t, \tilde{\delta}_t)$ and component parameters $(\Sigma_\mu, \Sigma_\delta)$, based on equations (3.2) and (3.3).
2: Based on the indicator variable γ, compute the regression component given the information about predictors at time $t + 1$ by equation (3.4).
3: Draw a random error in multivariate normal distribution with variance equal to Σ_ϵ and sum different components up in equation (3.1).
4: Sum up all the predictions and divide by the total number of MCMC iterations to generate the point prediction; establish the prediction intervals according to the corresponding quantile of predictors.

4 Main Findings and Model Performance Evaluation

In this section, we demonstrate our main findings and evaluate model performance. Specifically, in Section 4.1, the feature selection results generated by Algorithm 1 reveal that even

in the presence of both modern and classical predictors in both fundamental and technical sense, the polarity of news still adds on a complementary effect. In Section 4.2, the forecasts generated by Algorithm 2 on the test set indicate that sentimental predictors consistently boost the forecasting power. In this paper, we use the mean absolute forecast error (MAFE) to measure model performances. For Amazon and eBay, the in-sample MAFEs are 0.372 and 0.052, and the out-sample MAFEs are given in Table 4 of Section 4.3 as 22.84 and 0.451, which are both comparatively very small considering that the average stock prices are 851.487 and 31.365, respectively. Table 4 further reveals that all models under investigation with sentimental predictors outperform models without sentimental predictors, and the MBSTS model with sentimental predictors outperforms all the other models.

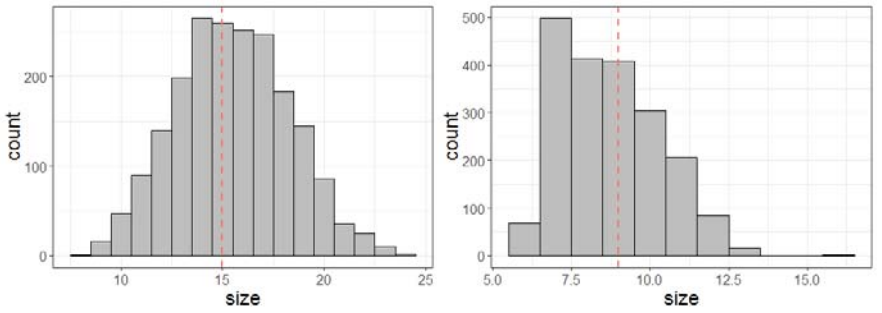
4.1 Importance of Financial News

News articles serve the purpose of spreading information about the companies, and further influence people either consciously or unconsciously in their decision-making process while trading in the stock market. Positive news such as good earnings reports, improved corporate governance, new products, and acquisitions, as well as positive overall economic and political indicators, translate into buying motivations and increasing in stock prices, while negative news will have opposite effects.

The "spike and slab" regression component of the MBSTS model enables feature selection and model training to be done simultaneously, which prevents overfitting and avoids redundant or spurious predictors. In this study, we run the model training Algorithm 1 for 2500 iterations, with the first 500 discarded as burn-in samples. All positive initial values for iterations were drawn from chi-square distributions, and others were drawn from normal distributions. It is worth noting that all predictors do not show an obvious trend, and most of them are stationary in the sense that their unit-root null hypotheses have p-values less than 0.05 in the augmented Dickey-Fuller test (see (Said and Dickey, 1984)). Therefore, we did not de-trend or de-season any predictors.

Although each company has 55 variables (27 Google domestic trends, 20 sentimental predictors, and 8 stock technical indicators) to choose from, from Figure 2 we can see that the median numbers of selected predictors are 15 for Amazon and 9 for eBay respectively, which is consistent with our prior setting of model sizes (more predictors are expected for Amazon). We can also see that although the model sizes around the median have a great number of counts, the range of model size is still large for both companies, which indicates model switching during MCMC iterations. In fact, although this is an undesirable property in other contexts, the model averaging technique incorporated in the MBSTS model can perfectly handle this by taking into account all the possibilities.

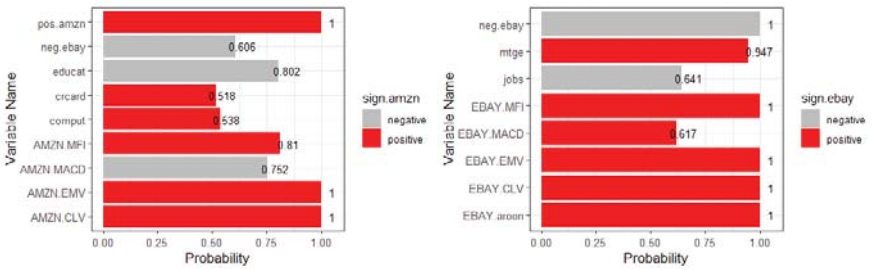
Figure 3 gives the features selected with posterior inclusion probabilities greater than 0.5. We can see that even in the presence of those successfully tested fundamental predictors and technical predictors, sentimental predictors were selected in all the model training experiments. However, we can also see that among all sentimental predictors, only positive news of Amazon (*pos.amzn*) and negative news of eBay (*neg.eBay*) significantly contribute to the variation in stock prices of these two companies, and no emotion score has a significant effect on the



(a) Amazon (b) eBay

Figure 2: Posterior distribution of model size

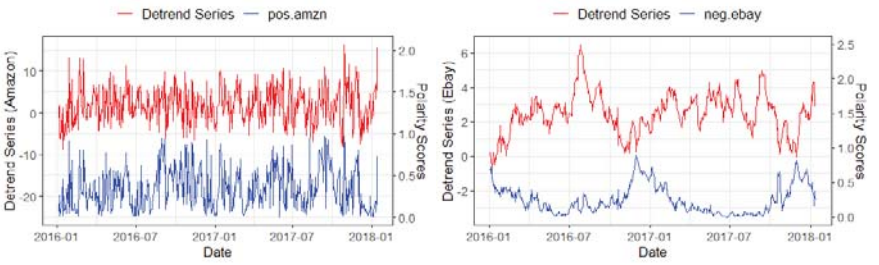
stock prices. It is easy to understand that positive news on Amazon pushes up its stock prices by boosting investors' confidence directly or indirectly, while negative news on eBay indicates current or potential unfavorable financial status which will have a negative effect on its stocks. The posterior inclusion probability of *neg.eBay* is 0.606 for Amazon, which can be explained by the fact that negative news on eBay will also negatively impact Amazon, since they both are in the e-commerce business and can face the same trouble.



(a) Amazon (b) eBay

Figure 3: Feature Selection by empirical posterior inclusion probabilities greater than 0.5. The bars are colored based on the sign of estimated regression coefficients.

To further investigate the prediction strength of financial news in explaining the fluctuations in stock prices movements, we compared polarity predictors *pos.amzn* and *neg.eBay* with the de-trend target series. In Figure 4(a), we can see that the polarity scores of positive news of Amazon shows very similar variation patterns to its de-trend time series; in Figure 4(b), we can see that the polarity scores of negative news of eBay do exhibit two perfectly matched peaks in the opposite direction of its de-trend time series. Considering the indicated sign of *neg.eBay* is negative (Figure 3), we can conclude that these two polarity predictors closely match the fluctuations where the local linear trend component is unable to capture. We can also conclude that even though there are other predictors with high posterior inclusion probabilities, these two sentimental predictors capture most of the sharp changes in the de-trended stock price time series.

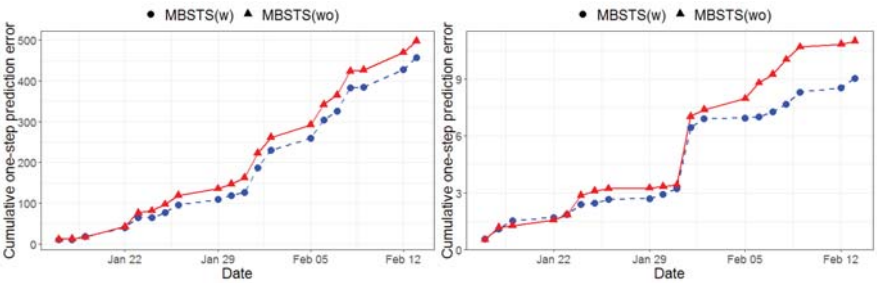


(a) Amazon (b) eBay

Figure 4: De-trend stock price series vs sentimental predictors with the highest marginal inclusion probabilities. (a) the correlation between de-trend series and pos.amzn is 0.726; (b) the correlation between de-trend series and neg.ebay is -0.497.

4.2 Forecast Errors on the Test Set

After investigating the influence of sentimental predictors on stock prices based on the training set, we explored their forecast performances on the test set in terms of cumulative one-step-ahead forecast errors. From Figure 5, we can see clearly that the MBSTS model with sentimental predictors beats its counterpart without sentimental predictors basically all the time and the accuracy advantage gets bigger as time goes by. That is to say, sentimental predictors consistently boost the forecasting power, which indicates its irreplaceable role in capturing the variations in stock prices that can not be explained by a linear trend component or other successfully tested predictors.



(a) Amazon (b) eBay

Figure 5: Cumulative one-step-ahead prediction errors in the year of 2018 by the MBSTS model with and without sentimental predictors.

Figure 6 presents a picture of true values and one-step-ahead prediction values with the help of sentimental predictors for both companies. It shows that all true values for each company fall into the 80% prediction band, which demonstrates strong prediction power of our methodology. In fact, besides the desired prediction accuracy, our approach can help predict the price values of a stock portfolio in after-hours trading, which refers to the transactions completed beyond regular trading hours including weekends and holidays through electronic communication net-

works (ECNs). During these times, polarity scores retrieved from breaking news if any, which usually has a huge impact on stock prices, can provide investors with guidelines on finding appropriate prices to ask or bid.

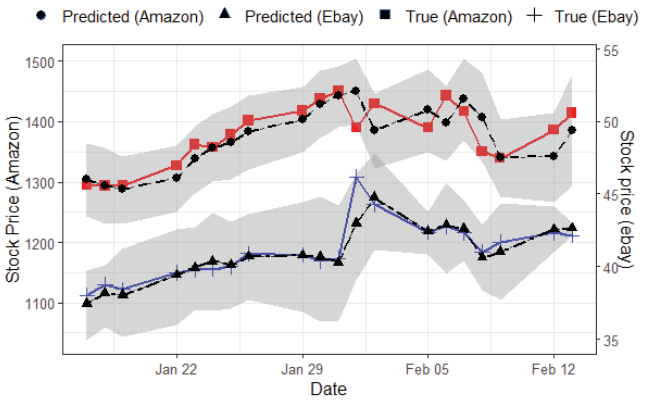


Figure 6: One-step-ahead predictions with 80% prediction interval over one month in the year of 2018

4.3 Model Comparison

In the following, we explicitly compare the prediction power contributed by news and public emotions in the framework of the MBSTS model, the ARIMA model, and the LSTM RNN model. The ARIMA model belongs to a class of models that capture a suite of different standard temporal structures in time series data. It incorporates the following three parts to better understand the data or to predict future points in the series: The autoregressive (AR) part indicates that the evolving variable of interest is regressed on its own lagged values; The moving average (MA) part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past; The integrated (I) part indicates that the data values have been replaced by the differences between their values and the previous values, where the initial differencing step may be applied one or more times to eliminate the non-stationarity. (Hyndman and Khandakar, 2008) proposed a stepwise algorithm, called *auto.arima*, that conducts a search over all possible models beginning with the selection of the I parameter, and then provide the best ARIMA model by minimizing the Akaike information criterion (AIC) to determine the values for the order of autoregressive terms and the order of the moving average process. In this study, we applied the widely used *auto.arima* algorithm to select the best fitted ARIMA model.

The LSTM RNN model is an RNN architecture that was designed to model temporal sequences and their long-range dependencies more accurately than conventional RNNs. It is well-suited to classifying, processing, and making predictions for the reason that there may be lags of unknown duration between important events in a time series. In this study, we also used the LSTM RNN forecast model for one-step univariate time series forecasting with and without predictors. For the LSTM RNN model, the batch size which is often much smaller than the total

number of samples, along with the number of epochs, defines how quickly the network learns the data, i.e. how often the weights are updated. Another important issue in defining the LSTM RNN layer is the number of neurons which also called the number of memory units or blocks, and a reasonably simple number between 1 and 5 should be sufficient. After hyper-parameter tuning, the following configurations with the minimum error were detected: Batch size equals to 1; The number of epochs equals to 3000; The number of neurons equals to 4. To reduce the influence of the initial setup on the model performances, we repeated the experiment 50 times, and then took the average of all one-step-ahead predicted values at each time point.

Model	Amazon		eBay	
	MAFE	Parameters	MAFE	Parameters
MBSTS(w)	22.84	55	0.451	55
MBSTS(wo)	24.87	35	0.509	35
LSTMX(w)	23.51	232	0.581	232
LSTMX(wo)	25.13	152	0.634	152
LSTM	26.48	12	0.814	12
ARIMAX(w)	25.48	57	0.652	56
ARIMAX(wo)	26.32	37	0.695	36
ARIMA	27.49	2	0.838	1

Table 4: Model comparison

Table 4 gives the prediction accuracy results on the test set in terms of MAFEs for the following models: the MBSTS model with sentimental predictors denoted as MBSTS(w), the MBSTS model without sentimental predictors denoted as MBSTS(wo), the LSTM RNN model with sentimental predictors denoted as LSTM(w), the LSTM RNN model without sentimental predictors denoted as LSTM(wo), the regular LSTM RNN model without any predictors denoted as LSTM, the ARIMA model with sentimental predictors denoted as ARIMA(w), the ARIMA model without sentimental predictors denoted as ARIMA(wo), the regular ARIMA model without any predictors denoted as ARIMA. We can see that MBSTS(w) has the smallest MAFE among all models for both companies. All models with sentimental predictors outperform models without sentimental predictors, and further outperform the standard models without any predictors.

Table 4 also provides the number of parameters that need to be learned through modeling training of all the models. MBSTS(wo) needs to learn 35 parameters for the 27 fundamental predictors and the 8 technical predictors. MBSTS(w) needs to learn additional 20 parameters for the 20 sentimental predictors. A common LSTM consists of 4 components (a cell, an input gate, an output gate, and a forget gate) and each component needs to learn 3 parameters (1 parameter for bias, 1 parameter for previous hidden state, 1 parameter for the lagged variable). For LSTMX(wo), each component has additional 35 parameters to learn for the 27 fundamental predictors and the 8 technical predictors, resulting in a total number of $(3 + 35) \times 4 = 152$ parameters. Further adding the 20 sentimental predictors, LSTMX(w) needs to learn $(3 + 35 + 20) \times 4 = 232$ parameters. The numbers of parameters to learn are the same for both Amazon and eBay of the above models, but are different when it comes to ARIMA. *auto.arima* indicates 1 AR parameter and 1 MA parameter for Amazon, while only 1

MA parameter for eBay. Correspondingly, adding the 27 fundamental predictors and 8 technical predictors, ARIMAX(wo) needs to estimate 37 parameters for Amazon and 36 parameters for eBay. Further adding the 20 sentimental predictors, ARIMAX(w) needs to estimate 57 parameters for Amazon and 56 parameters for eBay. From Table 4, we can see that MBSTS(w) has a remarkable performance with a moderate number of parameters to learn.

5 Concluding Remarks

In the framework of the MBSTS model, we creatively incorporated retrieved sentimental information from financial news and Twitter feeds for multivariate stock price time series prediction. Extensive experiments were conducted on two leading online commerce companies Amazon and eBay to examine whether sentimental predictors should be included, and further determine whether both polarity predictors and emotion predictors obtained through sentimental analysis have significant effects. The feature selection results reveal that even in the presence of other successfully tested predictors, polarity predictors should be included, while emotion predictors did not show similarly significant impacts on stock prices. The strength of polarity predictors in explaining the fluctuations that the local linear trend component could not, was confirmed, which is a special bonus.

Further analysis of the model performance with and without sentimental predictors on the test set in terms of cumulative one-step-ahead forecast errors, reveals that sentimental predictors consistently increased the forecasting power. From the one-step-ahead prediction results with 80% confidence interval over a month, we can clearly see that all the predictions capture the changing patterns and close to the actual values of stock prices. In a final comparison in terms of the mean absolute forecast error, to the traditional ARIMA model and the classical LSTM RNN model, as well as their with and without sentimental predictors counterparts, two conclusions emerge: First, it helps to include sentimental predictors; Second, the MBSTS model with sentimental predictors outperforms them all.

The model proposed in this paper is the first one that successfully embedded online text mining to an advanced multivariate Bayesian machine learning time series model, which opens the door of applying both text mining and machine learning simultaneously in modern quantitative finance research. However, we believe that the forecast power could still be possibly strengthened in three ways:

- Different models with better performances: In this paper, we compared the model performance to the standard and classical LSTM RNN model. However, nowadays, artificial neural network as a fast-growing and thriving field of artificial intelligence, has rapidly tackled and solved problems that are theoretical-unsolvable or theoretically-solvable-but-computationally-hard problems (see, for example, two wonderful new developments of advanced training and learning techniques regarding artificial neural networks (Kasabov, 2001) and (Ruiz-Rangel, Hernandez, Maradei Gonzalez and Molinares, 2018)). It is not hard to foresee that totally different models with better performances will be available in the future.

- Improved machine learning techniques: In this paper, the model under investigation consists of three main components: Kalman filter, Spike-and-slab method, and Bayesian model averaging, all of which work as a whole in the Bayesian paradigm to avoid overfitting. As new results coming out in regarding fields (see, for example, an excellent work on Bayesian filtering (Pozna, Precup, Tar, Škrjanc and Preitl, 2010)), we have enough reasons to believe that the model can be improved. Another possibility is to adopt another machine learning technique in time series prediction which is able to generate a better performance (see, for instance, a comparative study of different machine learning techniques used for forecasting (Azar, Moussa and Georges, 2018)).
- Advanced text mining techniques: In this paper, we used the simple lexicon-based approach for text mining, while advanced techniques could be used instead, such as algorithms for generating formal concepts and for constructing and navigating concept lattices (see, an in-depth analysis of multi-adjoint t-concept lattices (Medina and Ojeda-Aciego, 2010) and the references therein), which has wide applications in fields including data mining, text mining, machine learning, knowledge management, and semantic web.

Acknowledgment

We give special thanks to the journal editor and two anonymous reviewers for their constructive comments that led to significant improvement in the paper. Extensive code that the authors have written for fitting this model and data used in prediction, are available from the authors, on request.

References

Athanasiadis, I. N., Karatzas, K. D. and Mitkas, P. A. 2006. Classification techniques for air quality forecasting, *Fifth ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence, 17th European Conference on Artificial Intelligence, Riva del Garda, Italy*.

Azar, D., Moussa, R. and Georges, J. 2018. A comparative study of nine machine learning techniques used for the prediction of diseases, *International Journal of Artificial Intelligence* **16**(2): 25–40.

Banos, E., Katakis, I., Bassiliades, N., Tsoumakas, G. and Vlahavas, I. 2006. PersoNews: A personalized news reader enhanced by machine learning and semantic filtering, *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems", Montpellier, France*, Springer, pp. 975–982.

Bollen, J., Mao, H. and Zeng, X. 2011. Twitter mood predicts the stock market, *Journal of Computational Science* **2**(1): 1–8.

Bošnjak, M., Oliveira, E., Martins, J., Mendes Rodrigues, E. and Sarmento, L. 2012. Twitterecho: A distributed focused crawler to support open research with twitter data, *Proceedings*

of the 21st International Conference on World Wide Web, Lyon, France, ACM, pp. 1233–1240.

Carvalho, P., Sarmento, L., Silva, M. J. and De Oliveira, E. 2009. Clues for detecting irony in user-generated contents: Oh...!! It's so easy, *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion, Hong Kong, China*, ACM, pp. 53–56.

Dragoni, N., Gaspari, M. and Guidi, D. 2006. An infrastructure to support cooperation of knowledge-level agents on the semantic grid, *Applied Intelligence* **25**(2): 159.

Durbin, J. and Koopman, S. J. 2002. A simple and efficient simulation smoother for state space time series analysis, *Biometrika* **89**(3): 603–616.

George, E. I. and McCulloch, R. E. 1997. Approaches for Bayesian variable selection, *Statistica Sinica* pp. 339–373.

Glezakos, T. J., Tsiligiridis, T. A., Iliadis, L. S., Yialouris, C. P., Maris, F. P. and Ferentinos, K. P. 2009. Feature extraction for time-series data: An artificial neural network evolutionary training model for the management of mountainous watersheds, *Neurocomputing* **73**(1-3): 49–59.

Griffiths, W. E. 2003. Bayesian inference in the seemingly unrelated regressions model, *Computer-aided Econometrics*, CRC Press, pp. 263–290.

Harvey, A. C. 1990. *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. 1999. Bayesian model averaging: A tutorial, *Statistical Science* pp. 382–401.

Hyndman, R. J. and Khandakar, Y. 2008. Automatic time series forecasting: the forecast package for R, *Journal of Statistical Software* **27**(3).

Kasabov, N. 2001. Evolving fuzzy neural networks for supervised/unsupervised online knowledge-based learning, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **31**(6): 902–918.

Kasabov, N. and Song, Q. 2002. DENFIS: dynamic evolving neural-fuzzy inference system and its application for time series prediction, *IEEE Transactions on Fuzzy Systems* **10**(2): 144–154.

Katakis, I., Meditskos, G., Tsoumakas, G. and Bassiliades, N. 2009. On the combination of textual and semantic descriptions for automated semantic web service classification, *IFIP International Conference on Artificial Intelligence Applications and Innovations, Thessaloniki, Greece*, Springer, pp. 95–104.

Kontopoulos, E., Berberidis, C., Dergiades, T. and Bassiliades, N. 2013. Ontology-based sentiment analysis of twitter posts, *Expert Systems with Applications* **40**(10): 4065–4074.

Koutroumanidis, T., Iliadis, L. and Sylaios, G. K. 2006. Time-series modeling of fishery landings using ARIMA models and Fuzzy Expected Intervals software, *Environmental Modelling & Software* **21**(12): 1711–1721.

Laboreiro, G., Sarmento, L., Teixeira, J. and Oliveira, E. 2010. Tokenizing micro-blogging messages using a text classification approach, *Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data, Toronto, Canada*, ACM, pp. 81–88.

Lin, K.-P., Pai, P.-F., Lu, Y.-M. and Chang, P.-T. 2013. Revenue forecasting using a least-squares support vector regression model in a fuzzy environment, *Information Sciences* **220**: 196–209.

Madigan, D. and Raftery, A. E. 1994. Model selection and accounting for model uncertainty in graphical models using Occam's window, *Journal of the American Statistical Association* **89**(428): 1535–1546.

Medina, J. and Ojeda-Aciego, M. 2010. Multi-adjoint t-concept lattices, *Information Sciences* **180**(5): 712–725.

Motta, E., Domingue, J., Cabral, L. and Gaspari, M. 2003. IRS–II: A framework and infrastructure for semantic web services, *International Semantic Web Conference, Sanibel Island, FL, USA*, Springer, pp. 306–318.

Motta, E., Fensel, D., Gaspari, M. and Benjamins, R. 1999. Specifications of knowledge components for reuse, *Proceedings of SEKE '99, Kaiserslautern, Germany* pp. 36–43.

Musto, C., Semeraro, G. and Polignano, M. 2014. A comparison of lexicon-based approaches for sentiment analysis of microblog posts, *Information Filtering and Retrieval* **59**.

Pai, P.-F. and Lin, C.-S. 2005. A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega* **33**(6): 497–505.

Pai, P.-F., Lin, K.-P., Lin, C.-S. and Chang, P.-T. 2010. Time series forecasting by a seasonal support vector regression model, *Expert Systems with Applications* **37**(6): 4261–4265.

Pai, P.-F., Wei-Chiang, H., Ping-Teng, C. and Chen-Tung, C. 2006. The application of support vector machines to forecast tourist arrivals in Barbados: An empirical study, *International Journal of Management* **23**(2): 375–385.

Petris, G., Petrone, S. and Campagnoli, P. 2009. Dynamic linear models, *Dynamic Linear Models with R* pp. 31–84.

Pozna, C., Precup, R.-E., Tar, J. K., Škrjanc, I. and Preitl, S. 2010. New results in modelling derived from Bayesian filtering, *Knowledge-Based Systems* **23**(2): 182–194.

Preis, T., Moat, H. S. and Stanley, H. E. 2013. Quantifying trading behavior in financial markets using Google Trends, *Scientific Reports* **3**: 1684.

Qiu, J., Jammalamadaka, S. R. and Ning, N. 2018. Multivariate Bayesian structural time series model, *The Journal of Machine Learning Research* **19**(1): 2744–2776.

Ruiz-Rangel, J., Hernandez, C. J. A., Maradei Gonzalez, L. and Molinares, D. J. 2018. ERNEAD: Training of artificial neural networks based on a genetic algorithm and finite automata theory, *International Journal of Artificial Intelligence* **16**(1): 214–253.

Said, S. E. and Dickey, D. A. 1984. Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika* **71**(3): 599–607.

Sarmento, L., Carvalho, P., Silva, M. J. and De Oliveira, E. 2009. Automatic creation of a reference corpus for political opinion mining in user-generated content, *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion, Hong Kong, China*, ACM, pp. 29–36.

Scott, S. L. and Varian, H. R. 2014. Predicting the present with Bayesian structural time series, *International Journal of Mathematical Modelling and Numerical Optimisation* **5**(1-2): 4–23.

Scott, S. L. and Varian, H. R. 2015. Bayesian variable selection for nowcasting economic time series, *Economic Analysis of the Digital Economy*, University of Chicago Press, pp. 119–135.

Slini, T., Karatzas, K. and Moussiopoulos, N. 2002. Statistical analysis of environmental data as the basis of forecasting: An air quality application, *Science of the Total Environment* **288**(3): 227–237.

Slini, T., Karatzas, K. and Papadopoulos, A. 2002. Regression analysis and urban air quality forecasting: An application for the city of athens, *Global Nest* **4**(2-3): 153–162.

Strapparava, C. and Valitutti, A. 2004. Wordnet effect: an effective extension of Wordnet., *Proceedings of LREC, Lisbon, Portugal*, Vol. 4, Citeseer, pp. 1083–1086.

Tantrum, J., Murua, A. and Stuetzle, W. 2003. Assessment and pruning of hierarchical model based clustering, *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA*, ACM, pp. 197–205.

Tantrum, J., Murua, A. and Stuetzle, W. 2004. Hierarchical model-based clustering of large datasets through fractionation and refractionation, *Information Systems* **29**(4): 315–326.

Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A. and Kolehmainen, M. 2011. Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2. 5 concentrations using artificial neural networks, in Thessaloniki and Helsinki, *Science of the Total Environment* **409**(7): 1266–1276.

Wang, Z., Yan, W. and Oates, T. 2017. Time series classification from scratch with deep neural networks: A strong baseline, *2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, Alaska, USA*, IEEE, pp. 1578–1585.

Yan, W. 2012. Toward automatic time-series forecasting using neural networks, *IEEE Transactions on Neural Networks and Learning Systems* **23**(7): 1028–1039.

Yan, W., Qiu, H. and Xue, Y. 2009. Gaussian process for long-term time-series forecasting, *2009 International Joint Conference on Neural Networks, Atlanta, Georgia, USA*, IEEE, pp. 3420–3427.