Predicting Saudi Stock Market Index by Incorporating GDELT Using Multivariate Time Series Modelling

Rawan Alamro[1,2](B), Andrew McCarren[2](B), and Amal Al-Rasheed[1](B)

[1] Princess Nourah bint Abdulrahman University, Riyadh, Kingdom of Saudi Arabia
rawan.alamro3@mail.dcu.ie, aaalrasheed@pnu.edu.sa
[2] Dublin City University, Dublin, Ireland
andrew.mccarren@dcu.ie

Abstract. Prediction of financial and economic markets is very challenging but valuable for economists, business owners, and traders. Forecasting stock market prices depends on many factors, such as other markets' performance, economic state of a country, and others. In behavioral finance, people's emotions and opinions influence their transactional decisions and therefore the financial markets. The focus of this research is to predict the Saudi Stock Market Index by utilizing its previous values and the impact of people's sentiments on their financial decisions. Human emotions and opinions are directly influenced by media and news, which we incorporated by utilizing the Global Data on Events, Location, and Tone (GDELT) dataset by Google. GDELT is a collection of news from all over the world from different types of media such as TV, broad- casts, radio, newspapers, and websites. We extracted two time series from GDELT, filtered for Saudi Arabian news. The two time series rep- resent daily values of tone and social media attention. We studied the characteristics of the generated multivariate time series, then deployed and compared multiple multivariate models to predict the daily index of the Saudi stock market.

Keywords: Forecasting · Multivariate time series · Behavioral finance · Time series analysis

1 Introduction

People care a lot about the future: the future of their countries, families, finances, and relations. Predicting the future is considered a valuable skill nowadays. In the past, people dreamt of having this power. Ancient Greeks consulted the Oracle of Delphi, one of the most famous oracles of the Greek God Apollo. They consulted her knowing she was not infallible, and while they did not consider her revelations to be the objective truth, they still valued her visions. Interestingly, nowadays we take consultations from different experts and consultants with the

A. Alfaries et al. (Eds.): ICC 2019, CCIS 1097, pp. 317–328, 2019.
https://doi.org/10.1007/978-3-030-36365-9_26

same mentality of the ancient Greeks; we value their knowledge, but we do not believe in them blindly.

Experts use different types of metrics and data to predict a variety of measures. In the stock market, for example, economists use other markets to predict a single country's stock market movements. Other metrics used to predict stock markets are GDP, oil, financial freedom of a country, and many others. All previous predictors are based on reasoning and assume that people are logical creatures, which is not the case. People are emotional, and their emotions affect their decisions. Conventional economics assumes hard facts and rational decisions of people. This view has changed since the emergence of behavioural economics [1]. In behavioural economics, people's moods, emotions, and opinions have an impact on their transactional decisions and therefore on the financial markets. Nowadays, people are informed by different types of media such as news sites, news channels, forums, blogs, and even podcasts. Such vast amounts of infor- mation makes it very challenging to study the impact of news and information on financial markets. This leads us to the Global Data on Events, Location, and Tone (GDELT) dataset by Google. GDELT is a collection of news from all over the world from different types of media such as TV, broadcasts, radio, newspa- pers, and websites. Studying GDELT will provide a huge advantage to financial markets forecasters.

This research focuses on such effects on the Saudi Stock Market by studying the Tadawul All Share Index (TASI), which tracks the performance of all companies listed on the Saudi Stock Exchange. The research will use a time series analysis to predict the Saudi Stock Market Index by incorporating the GDELT dataset with the TASI.

2 Literature Review

The prediction of stock market trends is very important for the development of effective trading strategies [2]. Both statistical and machine-learning approaches are used to solve this type of problem and can provide traders with a reliable technique for predicting future prices [3,4]. In general, the prediction of the movement of stock prices is considered a difficult and important task for financial time series analysis. The accurate forecasting of stock prices is essential in assisting investors and traders to increase their stock returns. The natural noise and volatility in daily stock price fluctuation is the main reason for the high complexity in stock trends forecasting [5]. People opinions and emotions in media is showing promise for predicting financial markets. However, the true value of such data and which parameters can it predict is not agreed upon in the scien- tific society [1,6]. Modern financial theory is based on two main concepts: the efficient market hypothesis (EMH) and the capital asset pricing model (CAPM). Both concepts hold that investors can make rational responses to the market and that the stock market is unpredictable, ignoring the analysis of investors' actual decision-making behavior [1,6]. Since the development of behavioral finance, however, existing research shows that in the case of incomplete information, people's behavior, attitudes and preferences are not completely rational, that stock prices do not randomly fluctuate and, in some ways, the price is predictable. Not only will the news have an impact on the stock price, but so will the investors' mood [6]. Time series analysis and prediction are part of the wider field of data mining and analysis. A group of a large number of values within unified time interval is labelled as time series data. The time can be represented by year, month, week, day, etc. In time series analysis, the behavior and characteristics of a time series are studied and analyzed to predict future values and behavior, foreseeing the future data using the historical data in a timely structure [7,8].

2.1 Multivariate Time Series

Multivariate time series models are very valuable and popular in economics but much less so in other forecasting applications. Multivariate time series models, which are extensions to univariate ones, are given only a marginal position in standard textbooks on time series analysis. Outside economics, empirical examples are uncommon. In comparison, the multivariate perspective is fundamental in economics, where single factors are traditionally studied in the context of their relationships to other factors. Contrary to other fields of knowledge, economists may discard the use of univariate time series forecasting based on the interdependence theory, which appears to be an extravagant point of view [9].

Multivariate time series models are not necessarily better than univariate ones in forecasting, even in economics. While multivariate models are convenient in modeling fascinating interrelationships and achieving a better fit within a given sample, univariate methods are often found to outperform multivariate ones out of sample. And there are many possible reasons:

- There are more parameters for multivariate models than univariate ones. Each additional parameter is an unknown and has to be estimated. This estimate leads to an additional source of error due to variation in sampling.
- The number of potential multivariate model candidates exceeds their univariate counterpart. Therefore, model selection is more complex, takes longer time and more prone to errors, which affect prediction afterwards.
- Generalizing nonlinear algorithms to the multivariate models is difficult. In general, multivariate models need to be simple and to have basic structure compared to univariate models. This simplification is necessary to overcome the complication of multivariate models [9].

3 Methodology

This section discusses the methodologies used in this study to predict the Saudi Stock Market Index using a multivariate time series analysis. Integrating multivariate stochastic processes to be analyzed and modelled is a lengthy process that involves studying the time series characteristics and their integration and cointegration. We will use both statistical and machine learning approaches in the time series analysis to integrate GDELT dataset. Two time series from GDELT will be integrated to the model, which are Tone and Social Media Attention on Saudi Arabia. The Tone time series will represent a daily scale of the negativity and positivity of the news. While the Social Media Attention time series will represent how many news piece has been published daily concerning Saudi Arabia. The models chosen in this study are VAR, ARIMAX, multivariate GARCH, and LSTM, which have all proven successful in multivariate time series analyses and econometrics. Below is a generalized formula of the proposed models:

$$Y_t = aY_{(t-1)} + bZ_{(t-n)} + cX_{(t-m)} + e_t \tag{1}$$

where

Y_t = “stock market close price” at time t,
$Z_{(t-n)}$ = “Tone” at a previous lag n of time t,
$X_{(t-m)}$ = “Social Media Attention” at a previous lag m of time t,
e_t = error term at time t.

Figure 1 shows the methodology used in this study. Moreover, how the data is extracted from GDELT and Tadawul The Saudi Stock Exchange Company. We extracted both Tone and Social Media Attention time series from GDELT using Google’s BigQuery language in Google Cloud Platform, and TASI time series from Tadawul’s online portal. Finally, Fig. 1 represent the next phases of analysis, modeling and forecasting.

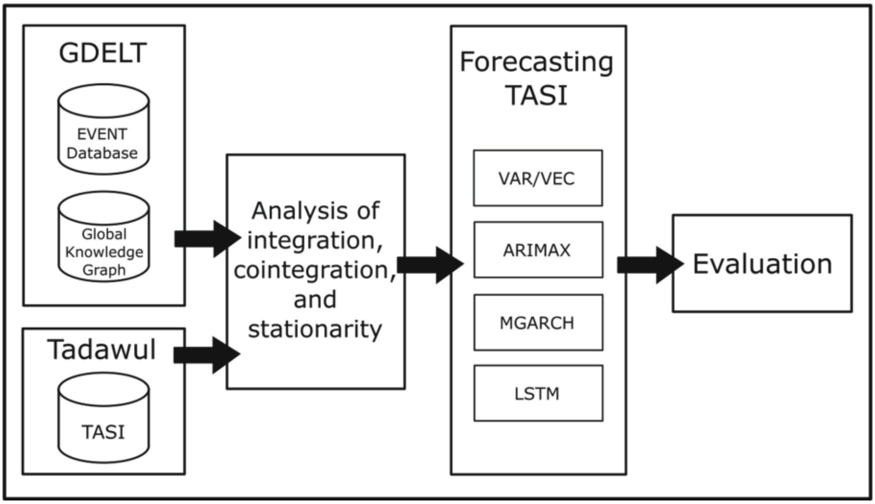

Fig. 1. Methodology graph.

3.1 Multivariate Time Series and Their Properties

Before explaining the methodology, we need to establish an understanding of multivariate time series and their special properties. Such understanding will give us a huge benefit in comprehending the data first and to examine any interrelationships. The following characteristics of multivariate time series are not all considered in the individual papers in the literature. Each research tries to cover one or two of them based on their choice of methodology. In this research, we try to cover all these characteristics to be able to apply it to the wide range of models used in the research.

Cross Correlation. One of the main objectives in multivariate time series analysis is to find relationships between the examined stochastic processes [12]. The study of cross correlation between time series help researchers to find any hidden relationships and connections. In the relationship between two time series X_t and Y_t, the series Y_t might be correlated to past lags of X_t. The Cross Correlation Function (CCF) is beneficial for detecting lags of X_t that might be usefull predictor of Y_t [10].

CCF is an important examining and assessment tool for modeling multivari- ate time series. The CCF generalizes the Autocorrelation Function (ACF) which is the autocorrelation function used on univariate time series to be used on the multivariate approach. Therefore, its main objective is to find linear dynamic correlations and interactions in time series data that have been produced from stationary processes [13]. Let X_t and Y_t represent a couple of stochastic processes that are stationary. Then the CCF is given as follows [13].

$$K_{XY}(T) = E[X_{t-T} Y_t] \tag{2}$$

Integration and Cointegration. In multivariate time series forecasting, cointegration is another important feature that need to be studied but usually neglected by beginners in statistics and econometrics. Cointegration examine non-stationary time series, which have variances, and means that differ over time. Therefore, cointegration allow researchers to identify the long-run rela- tionships or equilibrium in systems.

Integration. In statistics and econometrics, the order of integration I(d) of a time series is a statistical summary that state the minimum number of differences needed to convert the time series to a stationary one.

Cointegration. Cointegration is a statistical characteristic of multivariate time series or a collection of time series variables, which identify the degree to which these variables are sensitive to the same average over time. Therefore, confirm whether the distance between them remains constant over time. In the contrary, correlation return if the variables move to the same or opposite direction, which makes them positively or negatively correlated.

Two stochastic processes (time series) X_t and Y_t are called cointegrated if the following two conditions are fulfilled:

1. X_t and Y_t are both integrated processes of order one, i.e. X_t $I(1)$ and Y_t $I(1)$.
2. There exists a constant β /= 0 such that $Y_t - \beta X_t$ is a stationary process, i.e. $Y_t - \beta X_t$ $I(0)$.

The issue whether two integrated processes are cointegrated can be decided on the basis of a unit root test [3,13].

Endogenous Variables and Exogenous Variables. Endogenous and exogenous variables are widely used in econometrics and sometimes in statistics. Endogenous variables are similar to (but not the same as) dependent variables. Endogenous variables' values are influenced by other variables in the model. Meanwhile exogenous variable is not determined or influenced by any other variables in the model, although it could be affected by factors outside the model [3].

For example, a model trying to predict electricity consumption with the following available variables; electricity consumption, weather, financial situation of a country. The endogenous variables are electricity consumption, while the weather is considered an exogenous variable as it is not influenced by any other variable in the system. Financial situation is harder to classify as an exogenous variable, as we can argue that it could be influenced by weather and electricity consumption even in a subtle way. In this case, a variable can be considered partially exogenous and partially endogenous.

These characteristics of the variables are important in multivariate time series modeling, as some models are more appropriate to endogenous types such as VAR and some models are more appropriate to exogenous types such as ARI- MAX [3,10,15].

3.2 Vector Autoregressive and Vector Error Correction Models

In its basic form, a VAR consists of a set of K endogenous variables $y_t = (y_{1t},\ldots, y_{kt},\ldots, y_{Kt})$.

One important characteristic of a VAR(p)-process is its stability. This means that it generates stationary time series with time-invariant means, variances, and covariance structure, given sufficient starting values. One of the basic assumptions of VAR model is stationarity of all series, nonetheless differencing non-stationary time series individually considered a bad practice, as this might demolish important dynamic information. The critical issue is cointegration. In short, when there is cointegration, cointegrated models such as VEC (Vector Error Correction) should be used for forecasting. When there is no cointegration, series with an integrated appearance should be differenced [10,11].

The vector error correction (VEC) model is a special form of the VAR model which is used for variables that are stationary in their differences. VEC model is able to handle any cointegrating relationships between the variables and take it into consideration. VEC model is built with cointegration relations to restrict the long run effect of the endogenous variables to model their cointegrating relationship. The cointegration term is known in VEC as the error correction term as the deviation from long run equilibrium is corrected steadily through a series of short run amendments [10].

3.3 ARIMAX Model

Autoregressive Integrated Moving Average (ARIMA) models provide another approach to time series forecasting. ARIMA model is consider one of the widely used approaches in time series forecasting. ARIMA models aim to describe the autocorrelations in the data. ARIMA (p, d, q) is a combination of Autoregressive model AR(p) and a Moving Average model MA(q) for non-stationary time series where I(d) indicate the integration order.

Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) model extend ARIMA models through the inclusion of exogenous variables X. ARIMA is also an extension to another model, which is the Autoregressive Moving Average (ARMA) model. ARIMA model is used in some scenarios where data are non-stationarity, a preliminary differencing step can be implemented once or more to remove the non-stationarity [15,16].

3.4 Multivariate GARCH Model

The autoregressive conditional heteroscedasticity model (ARCH) is a statistical model for stochastic process that explain the variance of the current error term. Various problems in finance have motivated the study of the volatility or variability, of a time series. ARMA models were used to model the conditional mean of a process when the conditional variance was constant. However, the assumption of a constant conditional variance will be violated. Models such as the autoregressive conditionally heteroscedastic or ARCH model, first introduced by Engle [17] were developed to model changes in volatility. These models were later extended to generalized ARCH or GARCH models by Bollerslev [16]. The ARCH model is used traditionally when the error variance follows an autoregressive (AR) model. Therefore, the generalized autoregressive conditional heteroscedasticity (GARCH) model will be used if an autoregressive moving average (ARMA) model is assumed for the error variance, GARCH is a statistical model usually used in analyzing financial time series data [17].

Heteroscedasticity explain the irregular pattern of variation of a variable, or an error term in a statistical model. Essentially, when heteroscedasticity is assumed, values do not follow a linear pattern. In the contrary, they tend to cluster. Therefore, the results and predictive value one can extract from the model will not be reliable hence the necessity for using ARCH/GARCH models [18].

Although modelling volatility of the prices and returns was the main objective, understanding the co-movements of financial metrics is greatly important in practical cases. therefore, it is substantial to extend the GARCH model to the multivariate case (MGARCH) [19].

3.5 LSTM Model

Long Short Term Memory networks (LSTM) are a special kind of Recurrent Neural Network with the ability to learn long term relationships. They were introduced by Hochreiter & Schmidhuber [20], and were polished and distributed by many people in preceding work. They work very well on a many types of problems, and are now extensively used. All recurrent neural networks have the form of a looping chain of neural network. In traditional RNNs, this looping unit will have a very simple structure, for example; an individual tanh layer [21]. LSTMs are designed explicitly to avoid the problem of long-term dependency. Their default behavior is to remember information for long periods of time. LSTM is used in the deep learning field, contrary to standard forward neural networks. LSTM has feedback functionality that enable it to be a general purpose model. LSTM special properties enable it to be successful in many applications such as speech recognition, video processing, time series, and medical predictions [20]. LSTM model is well suited to forecast, process, and analyze time series data, as the duration between important events can be unknown, because of the different lags [6,22].

A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell [20].

4 Implementation

The dataset used in this study is GDELT developed by Google and TASI (Tadawul All Share Index), which is the Saudi stock market index. GDELT contains two main datasets; GDELT Event Database and GDELT Global Knowl- edge Graph (GKG). GDELT's entire quarter-billion-record database can be extracted using Google's BigQuery and Google Cloud Platform. The database is updated every 15 min and we can query, export, and analyze using SQL language [23].

Before forecasting, the data has been analyzed and pre-processed. The TASI data set represents the daily closing price of the Saudi market index. Media Attention represents the number of articles and new pieces across all media about Saudi Arabia, which is extracted from GDELT. Finally, Tone represents the tone of each of these news pieces from different media types, which is also extracted from GDELT. There was no missing values in the data. The Tone data has been smoothed to represent only one value per day; this aggregation has been completed using the mean Tone values for each day.

To analyze the time series, first we studied the cross correlation function to find any linear relationship between Tone, Media Attention, and TASI. The CCF shows that there is a strong negative relationship between TASI and SMA at lag 1 (as our data is daily, the lag is one day), and there is also a strong negative relationship between TASI and Tone at lag 1. These findings indicate that both Tone and Media Attention could have a predictive power over TASI [13,14] (Fig. 2).

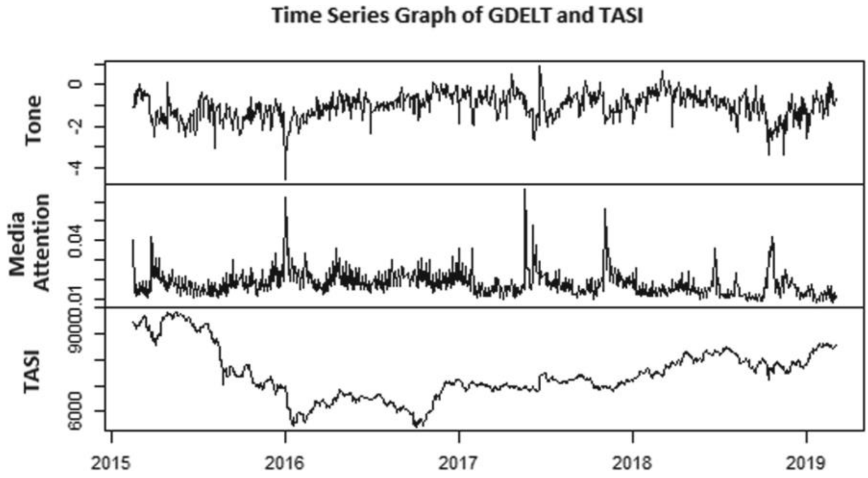

Fig. 2. Multivariate time series graph of GDELT and TASI.

Stationarity has been tested on all three time series using stationarity tests. Kwiatkowski-Phillips-Schmidt-Shin (KPSS), Augmented Dickey–Fuller (ADF), and Autocorrelation Function (ACF) stationarity tests have been conducted and showed that the TASI, Tone, and SMA are non-stationary and need to be differenced (first order) [10]. As all time series are non-stationary, we cannot model the data in the simplest form of VAR. In this case we also cannot difference the data before studying the cointegration of these time series, because if they are cointegrated some interrelation regressors might be lost in the process [10]. Johansen's test for cointegrating has been implemented and shows that there is cointegration between the series. Therefore we cannot use the simple form of VAR. The vector error correction (VEC) model has been used instead, which can handle the cointegration and non-stationarity.

After fitting the ARIMAX model, we studied the model residuals to ensure the good fit of the data. A residual analysis shows that the residuals have a con- stant mean, normally distributed, and goes gradually to zero in the ACF graph, which are all good indicators of a proper fit [15,16]. MGARCH was implemented using GO-MGARCH, which is one of the dynamic conditional correlation mod- els (DCC) in GARCH modelling. Multivariate GO-GARCH is able to explain and model the fluctuating volatility that is mostly present in stock market data and financial data [17]. In LSTM modelling, we split the data to train and test sets. After that, the input variables were reshaped into 3D format as restricted by LSTM (samples, time, and features) [22].

5 Evaluation

It is essential to evaluate forecast accuracy using actual forecasts. Therefore, the quantity of the residuals is not a reliable indication of the likelihood of the forecast errors. The forecast's accuracy can only be considered and therefore measured when the model is performed on new data that was not used in fit- ting the forecasting model. Traditionally, before modelling, researchers start by dividing the data into two parts, training data and testing data. Training data is used to estimate and generate the parameters of the model and the test data is used to calculate its accuracy. Because the test data is not used to estimate the model, it should be a reliable indicator of the model's predictive power on new data [11,24].

To evaluate the different models implemented for forecasting, we used error measures and scale-dependent measures. The two most commonly used scale-dependent measures are based on the absolute errors or squared errors [24]:

$$Meanabsoluteerro : MAE = mean(|e_t|) \tag{3}$$

$$Rootmeansquarederror : RMSE = \sqrt{(mean(e_t^2))} \tag{4}$$

In forecasting time series, error measures are the most commonly used accuracy measures. A study forecasting crude oil prices based on internet concern used MAE, RMSE, and MAPE as error measures for the prediction [24]. Another study used MAE and RMSE to evaluate the prediction of crude oil prices based on deep learning modelling using text [4].

MAE is more popular than RMSE, as it is easier to implement and under- stand. The table below shows the results of each model. We can conclude that LSTM is the most accurate model, scoring 0.59 MAE, and that VAR and mul- tivariate GARCH are the lowest performing models (Table 1).

Table 1. Forecasting results.

Model\error measure	RMSE	MAE
VAR\VEC	8080.49	8074.12
ARIMAX	325.06	284.75
Multivariate GARCH	8038.9	8032.5
LSTM	0.61	0.59

The LSTM loss function demonstrate that the training performance and the test performance were almost the same after the thirtieth iteration, which proves that the model is not overfitting.

The results shows that the VAR model performance is low compared to LSTM and ARIMAX, which can be explained by the endogenous/exogenous characteristics of the variables. The VAR model assumes all variables are endogenous. In our model, the TASI variable, which is the stock market index closing price, is endogenous, as it depends on the other two variables, Media Attention and Tone. Meanwhile, Media Attention and Tone can be partially endogenous and partially exogenous, as we are not certain. We can argue that the stock market index could affect Media Attention and Tone but not strongly, which makes them more exogenous, hence the success of ARIMAX compared to VAR [3].

A disadvantage of the multivariate GARCH model is that the number of parameters to be estimated in the equation increases rapidly, which consequently leads to a higher degree of errors, hence the high error measure of the multivariate GARCH in this research [19].

6 Conclusion

Predicting stock market movements is critical for the development of effective trading tactics. Both statistical and machine learning approaches are used to solve this problem and can provide traders a reliable technique for predicting future prices. Forecasting stock market prices is a challenging task, as stock markets depend on multiple factors that vary in type and extraction complexity. Financial markets can be influenced by economic factors and non-economic fac- tors, which are harder to figure and to analyse. Human behaviours, emotions, and sentiments are an example of non-economic factors that are difficult to track and extract. In this study, we try to solve this issue by introducing GDELT as a global media dataset that has different useful features. We extracted two time series from GDELT, filtered for Saudi Arabian news. The two time series represent daily values of tone and social media attention. We studied the characteristics of the generated multivariate time series, then deployed and compared multiple multivariate models to predict the daily index of the Saudi stock market. The results show that the model with the highest performance is LSTM, with 0.59 MAE. This concludes that LSTM can give very accurate forecasts, as it has a very low MAE compared to the mean value of TASI, which is 7413.

References

1. Ogaki, M., Tanaka, S.C.: Behavioral Economics. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-6439-5
2. Elshendy, M., Fronzetti Colladon, A.: Big data analysis of economic news: hints to forecast macroeconomic indicators. Int. J. Eng. Bus. Manag. **9**, 1–12 (2017)
3. Mills, T.C.: Time Series Econometrics. Springer, Cham (2015). https://doi.org/10.1057/9781137525338
4. Li, X., Shang, W., Wang, S.: Text-based crude oil price forecasting: a deep learning approach. Int. J. Forecast. **35**, 1548–1560 (2018)
5. Ticknor, J.L.: A Bayesian regularized artificial neural network for stock market forecasting. Expert Syst. Appl. **40**(14), 5501–5506 (2013)
6. Zhang, G., Xu, L., Xue, Y.: Model and forecast stock market behavior integrating investor sentiment analysis and transaction data. Cluster Comput. **20**(1), 789–803 (2017)
7. Fakhrazari, A., Vakilzadian, H.: A survey on time series data mining. In: IEEE International Conference on Electro Information Technology, pp. 476–481 (2017)

8. Esling, P., Agon, C.: Time-series data mining. ACM Comput. Surv. **45**(1), 1–34 (2012)
9. Kunst, R.M., Franses, P.H.: The impact of seasonal constants on forecasting seasonally cointegrated time series. J. Forecast. **17**(2), 109–124 (1998)
10. Pfaff, B.: Analysis of Integrated and Cointegrated Time Series with R. Springer, New York (2008). https://doi.org/10.1007/978-0-387-75967-8
11. Shanmugam, R., Brockwell, P.J., Davis, R.A.: Introduction to time series and forecasting. Technometrics **39**(4), 426 (1997)
12. Kunst, R.: Multivariate forecasting methods. Building 28–39 (2000). https://homepage.univie.ac.at/robert.kunst/prognos4.pdf
13. Beran, J.: Mathematical Foundations of Time Series Analysis: A Concise Introduction. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74380-6
14. Hunter, J., Burke, S.P., Canepa, A.: Multivariate Modelling of Non-Stationary Economic Time Series. Palgrave Macmillan, London (2005)
15. Elshendy, M., Colladon, A.F., Battistoni, E., Gloor, P.A.: Using four different online media sources to forecast the crude oil price. J. Inf. Sci. **44**(3), 408–421 (2018)
16. De Gooijer, J.G., Hyndman, R.J.: 25 years of time series forecasting. Int. J. Forecast. **22**(3), 443–473 (2006)
17. Wenjing, S., Huang, Y.: Comparison of multivariate garch models with application to zero-coupon bond volatility. Thesis, pp. 1–55 (2010)
18. Kartsonakis Mademlis, D., Dritsakis, N.: Volatility between oil prices and stock returns of dow jones index: a bivariate GARCH (BEKK) approach. In: Tsounis, N., Vlachvei, A. (eds.) ICOAE 2018. SPBE, pp. 209–221. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02194-8 16
19. Bauwens, L., Laurent, S., Rombouts, J.V.K.: Multivariate GARCH models: a survey. J. Appl. Econom. **21**(1), 79–109 (2006)
20. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**, 1735–1780 (1997)
21. Greff, K., Srivastava, R.K., Koutnik, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: a search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. **28**(10), 2222–2232 (2017)
22. Fischer, T., Krauss, C.: Deep learning with long short-term memory networks for financial market predictions. Eur. J. Oper. Res. **270**(2), 654–669 (2018)
23. Leetaru, K., Schrodt, P.A.: GDELT: global data on events, location and tone. International Studies Association (2012)
24. Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. Int. J. Forecast. **22**(4), 679–688 (2006)