International Review of Economics and Finance 103 (2025) 104403

Contents lists available at ScienceDirect

International Review of Economics and Finance

journal homepage: www.elsevier.com/locate/iref

Value-at-risk forecasting- based on textual information and a hybrid deep learning-based approach

Yangfan Cao[a,*], Wei Chong Choo[b], Bolaji Tunde Matemilola[c]

[a] School of Business School, Sunway University, Sunway City, Selangor, Malaysia
[b] Department of Management and Marketing, Faculty of Economics and Management, Universiti Putra Malaysia, UPM, Serdang, Selangor, 43400, Malaysia
[c] Department of Accounting and Finance,Faculty of Economics and Management, Universiti Putra Malaysia, UPM, Serdang, Selangor, 43400, Malaysia

ARTICLE INFO

Keywords:
VaR
GARCH
LSTM
CNN
NLP
Sentiment analysis

ABSTRACT

The recent rise in deep learning and natural language processing (NLP) applications has notably improved productivity across different fields. This research aims to refine Value-at-Risk (VaR) model accuracy by leveraging text mining and deep learning. It first uses NLP to analyze online news sentiments, integrating these as variables to boost stock market risk forecasts and assess their effect on VaR accuracy. Additionally, the study combines predictions from four unique Generalized AutoRegressive Conditional Heteroskedasticity (GARCH)-type models into advanced Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN)-LSTM models to see if this boosts VaR precision. It also explores how textual data impacts VaR predictions over short and longer periods, using 7 and 20-day rolling windows. The analysis, using S&P500 (SPY), Dow Jones Industrial Average (DJI), and Nasdaq Composite (IXIC) data from 2012 to 2023 alongside news headlines, tests these approaches. The results confirm that incorporating textual information into the VaR model enhances its forecasting accuracy, highlighting the benefits of applying deep learning techniques in this process.

1. Introduction

The Value-at-Risk (VaR) is extensively utilized as an indicator of financial market risk, serving both regulatory and internal risk management purposes. Essentially, VaR is a straightforward statistical concept and a pragmatic tool for assessing risk. It quantifies the maximum potential loss within a specified time and with a defined confidence level. Beyond its crucial role in risk management across diverse sectors, where it has gained widespread acceptance as the primary risk metric, VaR also serves as a significant factor in portfolio allocation (Giacometti et al., 2007). Particularly, the global economic downturn triggered by the onset of the COVID-19 pandemic has raised heightened concerns about risk management. Accurate control and prediction of financial risk are crucial for informed decision-making by key stakeholders, including managers and regulators in risk management (Meng and Taylor, 2019). While the conceptual framework of VaR is clear, it is imperative to recognize that achieving precise measurement of conditional VaR poses statistical challenges (Bali et al., 2008). Furthermore, Abad and Benito (2013) underscore the substantial influence of volatility models on the precision of VaR forecasting. In essence, the precision of VaR estimates can be significantly improved by employing

* Corresponding author.
E-mail address: caoyangfanchris@outlook.com (Y. Cao).

https://doi.org/10.1016/j.iref.2025.104403
Received 5 February 2025; Received in revised form 21 May 2025; Accepted 4 July 2025
Available online 7 July 2025

accurate volatility models. Conventional econometric approaches have a key strength in their ability to be theoretically illustrated through statistical principles. Nonetheless, these methods operate under the assumption that explanatory variables must invariably be stationary, and their performance prominently diminishes when dealing with a combination of qualitative variables (Kim & Won, 2018). Furthermore, the utilization of qualitative data presents challenges, as it differs from quantitative data, which is acquired and organized based on mechanistic comprehension (Li, 2019).

With the emergence of text mining technology, text mining methods have primarily been employed to identify opinions and extract information. Studies indicate that sentiment indicators derived from headlines have a notable impact on stock markets. Figà-Talamanca and Patacca (2024) found that a Twitter-based sentiment index influences stock returns, while a news-based index affects market volatility, emphasizing the importance of concise information in shaping investor behavior. Increasingly, research has demonstrated that leveraging sentiment analysis for quantifying news headlines can significantly simplify the utilization of qualitative data and improve volatility forecasting (Figà-Talamanca & Patacca, 2024; Jiao et al., 2022; Li et al., 2019). Moreover, the stock market stands out as one of the most responsive domains to people's sentiments, which have the potential to influence overall market trends. Certainly, market sentiment has been a longstanding focus in stock markets, often alongside fundamental and technical analysis. However, there's a scarcity of studies dedicated to transforming qualitative data, like sentiment and subjectivity information, into quantitative data for the purpose of constructing features in VaR models to enhance prediction accuracy. Consequently, this study employs natural language processing (NLP) techniques to extract sentiment and subjectivity components from news headlines, integrating them as independent variables. This approach is adopted to enrich the model architecture with more qualitative information.

In a prior study, Li et al. (2019) indicate that crude oil price forecasting models proposed through a feature grouping method based on the Latent Dirichlet Allocation (LDA) topic model outperform older benchmark models. Consequently, this investigation employs the LDA topic model to uncover hidden topics in online news, aiming to provide richer textual information for the construction of VaR models. The objective is to assess whether latent topics extracted from online news headlines can enhance the forecasting performance of VaR models. Upon examining our news data, we identified several overarching topics for stock market news headlines, including 'Market Movements', 'Market Trends and China Impact', 'Dow, European, and Data Impact', 'Earnings Impact', and others. Each of these topics potentially links to a factor influencing VaR predictions. This study further examines how these identified topics, treated as independent variables, influence our VaR estimates.

Nevertheless, a notable concern arises in the form of multicollinearity when utilizing topics generated by LDA as independent variables in regression. Slof et al. (2021) observed this issue and suggested discarding some LDA-generated topics to mitigate multicollinearity. Addressing this challenge more effectively, Chan et al. (2022) demonstrated that machine learning algorithms outperform simple OLS estimators when dealing with data afflicted by multicollinearity. Not requiring information about data relationships or distributions, these algorithms, particularly deep learning models, are deemed suitable for application in this study to avoid multicollinearity problems. Besides, investor sentiments can be shaped by varying investment horizons and rolling periods (Bouteska et al., 2023). Thus, our objective is to assess the accuracy of VaR estimates under the influence of different rolling windows. This involves a comparison between two specific rolling periods, with an additional evaluation of how textual data influences model performance within these distinct rolling windows.

Furthermore, due to the autocorrelation and nonlinearity inherent in financial assets' return series, which can compromise model performance, accurately modeling the volatility of asset prices has historically been challenging (Kakade et al., 2022). This study introduces VaR forecasting models that integrate information from various GARCH-type models, such as GARCH, IGARCH, EGARCH, and GJRGARCH, which are commonly used statistical models to characterize diverse features in financial time series. These models are incorporated into a hybrid Long short-term memory (LSTM) neural network, effectively addressing the issues associated with volatility clustering in the data through the GARCH models and their variants, while leveraging the LSTM's capabilities in modeling the nonlinear and long-memory aspects of asset price volatility series.

Convolutional neural network (CNN) models excel in filtering input data noise and extracting valuable features, making them a favored and effective deep learning approach widely employed in text processing and image recognition (Livieris et al., 2020). Moreover, the CNN-LSTM combination has proven effective as a time-series data model across various applications (Choi and Shin, 2022; Vidal and Kristjanpoller, 2020). In fact, limited research has focused on the precise prediction of VaR using CNN-LSTM. Additionally, a support vector machine (SVM) model with asset-specific input features (news sentiment and asset-specific data) ranked first among their respective model classifiers, suggesting that both Bitcoin news sentiment and asset-specific information are valuable factors in predicting the next day's price direction (Gurrib & Kamalov, 2022). Koukaras et al. (2022) integrated multiple sentiment analysis and machine learning methods, emphasizing the extraction of additional features from social media to improve stock prediction accuracy. Jiao et al. (2022) incorporated sentiment indicators into deep learning for predicting financial asset volatility, significantly enhancing the accuracy of crude oil volatility forecasting. While previous research has verified the superior integration of machine learning models and sentiment information, limited studies have focused on applying this approach to risk forecasting. The study evaluates the performance of the proposed models against conventional ones using the S&P 500 (SPY), Dow Jones Industrial Average (DJI), and Nasdaq Composite (IXIC), the representative American stock index. The collected textual data from associated stock market news headlines undergo NLP to convert qualitative data into quantitative data. Employing CNN layers enhances feature extraction for LSTM construction, contributing to more accurate VaR forecasting.

Section 2 provides a concise overview of the methodology, divided into three main parts. The first and second parts detail the construction of the quantitative and qualitative components. The remainder of Section 2 examines the literature on financial time-series models, deep learning techniques, and explains how quantitative and qualitative information is integrated. Section 3 presents quantitative and qualitative information regarding the experimental procedures, evaluation methods, outcomes, and analysis of model topics. The empirical study and data analysis are expounded in Section 4. The paper concludes in Section 5, offering recommendations for further research.

2. Methodology

2.1. Qualitative indicators construction

2.1.1. Sentiment analysis

For better predicting the VaR of SPY, DJI, and IXIC the textual variables and some quantitative variables are included that have a positive impact on VaR estimation in this study. After collecting textual data from news headlines, preprocessing steps such as tokenization, lowercasing, stop word removal, stemming, and lemmatization are applied. From these processed headlines, two types of latent quantitative features, namely emotional polarity scores and subjectivity scores, are extracted for use in feature construction. In previous study, Jiao et al. (2022) suggested incorporating sentiment indicators to predict financial asset volatility, highlighting the significant influence of investor sentiment on volatility. Similarly, Deveikyte et al. (2022) found a strong correlation between social media sentiment (such as Twitter mood) and next day volatility. These results are consistent with Jiao and Walther (2020), who observed that heightened social media activity related to a particular company is associated with a notable rise in return volatility. Overall, there is growing evidence supporting the use of investor sentiment measures such as news headlines, social media content, and sentiment indices as qualitative inputs in volatility modeling. These additions aim to enrich information diversity with contextual sentiment insights while addressing the inherent complexity of volatility, ultimately enabling hybrid models to better capture the nonlinear dynamics (e.g., abrupt sentiment-driven regime shifts) and long-memory dependencies (e.g., prolonged effects of market mood fluctuations) observed in asset price volatility series.

This research also utilizes the Textblob Python library, offering a straightforward API for engaging in various natural language processing tasks, including sentiment analysis, classification, and translation. The sentiment module of Textblob furnishes two scores for daily news headlines: polarity and subjectivity. Polarity scores fall within the range of $[-1.0, 1.0]$, where values below zero denote negative news, above zero indicate positive news, and zero represents neutral news. Similarly, subjectivity scores range from $[0.0, 1.0]$, with 0.0 signifying high objectivity and 1.0 indicating high subjectivity.

2.1.2. Latent Dirichlet Allocation topic model

We employ the LDA modeling approach developed by Blei et al. (2003) to unveil latent topics within the corpus of news headlines. The implementation of the LDA model is facilitated using the Python library gensim. The underlying assumption of the LDA model is that each document is a blend of multiple topics, with each topic having a corresponding probability distribution for different words. However, obtaining precise inferences for various topic distributions and word probabilities across each topic in this model is generally computationally challenging. Based on our empirical findings, the textual features derived from our study encompass valuable predictive information for the volatility of three stock indices. The LDA model indicates that the distinctive information embedded in news text primarily pertains to stock market movements, the relationship of the US market with other countries or regions, and trends in stock trading. The outcomes from LDA in this study represent the degree of relevance of each topic within daily news headlines, which are employed as independent variables for model construction. Additionally, further details and results of the LDA will be presented in the next section. Notably, integrating such textual information from news with financial market data in machine learning-based forecasting models leads to substantial enhancements in accuracy.

2.2. Quantitative indicators construction

Quantitative indicators derived from historical price and volume data have proven to be highly effective in financial forecasting. For instance, Neely, Rapach et al. (2014) employed such indicators to directly predict stock returns. Liu and Pan (2020) demonstrated that incorporating a comprehensive set of technical variables—based on price, volume, and volatility—into an autoregressive model significantly improves the accuracy of volatility forecasts compared to models that exclude them. Similarly, Jiao et al. (2022) applied price movement and risk factor classification within an LSTM framework for volatility prediction, finding that price movement, in particular, serves as a crucial independent variable for enhancing forecast accuracy. Specifically, the risk factor classification is based on whether the previous volatility value exceeds a certain threshold value, and the price movement is based on whether the previous day's price rises or falls when compared to the price from two days prior.

The following is the definition of the formula for price movement based on price variation:

$$S_{t-1} = \begin{cases} 0, \; p_{t-2} > p_{t-1} \\ 1, \; p_{t-2} < p_{t-1} \end{cases}$$

where p_{t-1} represents the closing price of the stock index on previous day and p_{t-2} represents the closing price of the stock index on the previous 2 days. Risk factor classification is constructed based on whether the day's volatility value exceeds a setting threshold value.

$$W_{t-1} = \begin{cases} 0, V_{t-1} < V \\ 1, V_{t-1} > V \end{cases}$$

where V_{t-1} represents the volatility of the U.S. stock indices (DJI, IXIC, and SPY) on previous day and V represents the risk threshold established by the upper four quartiles of the unconditional volatility indices for DJI, IXIC, and SPY. The excess of the risk threshold V indicates that the U.S. stock market has certain risks. For more technical details, one can refer to Li et al. (2019).

Besides, the emotion index active ratio (AR) and willingness ratio (BR) are the common indicators of stock analysis and will be utilized as the independent variables in this study since they display the stock market's popularity and willingness of the market to buy and sell. The calculation formula is as follows:

AR: (closing price - opening price)/(opening price - lowest price)*100

BR: (highest price - previous closing price)/(previous closing price - lowest price)*100.

In addition, the GARCH model for volatility forecasting is beneficial from trading volume as an external regressor. Fuertes et al. (2015) found that integrating daily trading volume with return data leads to the greatest economic benefits in volatility forecasting. Building on this, the present study incorporates two volume based variables originally proposed by Liu et al. (2020) as part of the quantitative indicators used for predicting daily volatility. Particularly, the two alternative trading volume quantitative variables are the natural logarithm of lagged volume, and also the high/low volume indicator variable, $IndVol_{t-1}$, as shown:

$$IndVol_{t-1} = \begin{cases} 1 \ \ if \ IndVol_{t-1} \geq \dfrac{1}{n-1}\sum_{i=2}^{n} Volume_{t-i} \\ 0 \ \ if \ IndVol_{t-1} < \dfrac{1}{n-1}\sum_{i=2}^{n} Volume_{t-i} \end{cases}$$

2.3. Financial time-series models (Historical Volatility indicators)

2.3.1. Value-at-risk

Modern Portfolio Theory (MPT) identifies variance, or volatility, as a fundamental measure of risk in portfolio selection. Proposed by Markowitz (1952), this theory formalized the trade-off between risk and return by quantifying expected return with the mean and risk with the variance. This framework laid the foundation for contemporary risk management practices in the financial sector and inspired the development of various risk measures, including VaR. Today, VaR forecasting methods are generally categorized into three main types: parametric, nonparametric, and semiparametric approaches. There are no presumptions regarding the distribution of returns or the parametrization of conditional volatility in nonparametric approaches. In earlier research (Kakade et al., 2022; Lu et al., 2022; Zhang et al., 2022), parametric methods are used for VaR forecasting because volatility forecasting is a necessary step in the process when deep learning models are implemented into VaR forecasting. A VaR estimate is obtained via the parametric approach by:

$$VaR_t = \widehat{\mu}_t + F_\alpha \widehat{\sigma}_t \tag{1}$$

The conditional mean and standard deviation are denoted by $\widehat{\mu}_t$ and $\widehat{\sigma}_t$, and F_α is the crucial value established by the confidence interval and the selected theoretical distribution. In this study, predicted volatility from various GARCH models is utilized as the independent variable to build deep learning models. These deep learning models integrate information from different GARCH-type models to enhance VaR forecasts. The properties of several conditional variance models along with the conditional mean model will be discussed in the remaining sections.

1) ARMA

Assume that the independent random variables $r_1, r_2, r_3, ..., r_t$ are all equally distributed and that $r_t = log(P_t) - log(P_{t-1})$ reflects the asset return from time t - 1 to time t. The models we'll contrast depend on r_t carrying out the actions listed below (Engle & Patton, 2001):

$$r_t = \mu_t + \sigma_t \in_t, \in_t \sim \ i.i.d. \ F(\cdot) \tag{2}$$

Where F is the cumulative distribution function of $\in_t$, conditional mean and conditional volatility are expressed as μ_t and σ_t, respectively. A complete model for a parametric model should explain how μ_t and σ_t are dependent on the historical data as well as the distribution F. The mean process specification has no appreciable influence on VaR estimation, according to Angelidis et al. (2004). In order to represent the conditional mean r_t for building VaR, a linear autoregressive moving average ARMA (p, q) model is used throughout the paper:

$$\widehat{r}_t = \alpha_0 + \sum_{i=1}^{p} \beta_1 r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta \varepsilon_{t-j} \tag{3}$$

Where ε_t is calculated as the variance of an error term using the formula: $\varepsilon_t = r_t$ - μ_t. The price "new" or "shock" are common terms used to describe it.

2) GARCH

Engle (1982) first introduced ARCH models, that demonstrate conditional variance as a linear function of lagged squared error components. The issue that it is challenging to establish the order (p) of ARCH(p) is resolved by the GARCH models, which Bollerslev (1986) developed as an extension of the ARCH models. In GARCH models, conditional variance is expressed as a linear function using lagged squared error terms as well as lagged conditional variance terms. The GARCH (p, q) model is used as an example and is presented by:

$$\widehat{\sigma}_t^2 = \omega + \sum_{i=1}^{m} \alpha_1 \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_1 \sigma_{t-j}^2 \tag{4}$$

Where $\omega > 0$, both the coefficients α_1 *and* β_1 are non-negative, resulting in a positive and stationary predictive variance.

3) IGARCH

To explore volatility persistency, Engle and Bollerslev (1986) developed the IGARCH model, which is the unit-root GARCH model. Similar to ARIMA model, a key feature of IGARCH model is that the impact of past squared shocks $\eta_{t-i} = \varepsilon_{t-i}^2 - \sigma_{t-i}^2$ for $i > 0$ on ε_t^2 is persistent. IGARCH(p, q) models can be expressed as follows:

$$\widehat{\sigma}_t^2 = \omega + \sum_{i=1}^{m} (1-\beta_1)\varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_1 \sigma_{t-j}^2 \tag{5}$$

Where $1 > \beta_1 > 0$, it guarantees the variance process' stationarity and non-negativity.

4) EGARCH and GJRGARCH

Studies of financial stock returns typically demonstrate that periods of greater volatility follow periods of negative returns more so than periods of positive returns of equivalent size. This imbalance can be explained by the fact that a company's leverage is affected differentially by positive and negative shocks, and this in turn impacts how volatile it is (Black, 1976). This leverage effect has been addressed by introducing more adaptive volatility treatments that take into account the asymmetric volatility induced by positive and negative shocks. These models are referred to as non-linear since the conditional variance is no longer treated as a linear function of lagged squared error and lagged variance. Two models that capture the asymmetric nature of returns are the exponential GARCH (EGARCH) model (Nelson, 1991) and the Glosten-Jagannathan-Runkle GARCH (GJRGARCH) model (Glosten et al., 1993). The following describes the conditional variance EGARCH (1, 1) model:

$$In(\widehat{\sigma}_t^2) = \omega + \alpha_1 \frac{|\varepsilon_{t-1}| + \gamma_1 \varepsilon_{t-1}}{\sigma_{t-1}} + \beta_1\, ln(\sigma_{t-1}^2) \tag{6}$$

When ε_{t-1} is positive, it increases $\alpha_1(1+\gamma_1)|\,\varepsilon_{t-1}/\sigma_{t-1}|$ to the log volatility, while when ε_{t-1} is negative, it adds $\alpha_1(1-\gamma_1)|\,\varepsilon_{t-1}/\sigma_{t-1}|$.

Another general asymmetric GJRGARCH (1, 1) is specified as:

$$\widehat{\sigma}_t^2 = \omega + (\alpha_1 + \gamma_1 N_{t-1})\varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{7}$$

Where N_{t-1} is an indicator function, that is,

$$N_{t-1} = \begin{cases} 1 \;\; if\; \varepsilon_{t-1} < 0 \\ 0 \;\; if\; \varepsilon_{t-1} \geq 0 \end{cases} \tag{8}$$

And α_1, γ_1 and β_1 are non-negative coefficients satisfying conditions like those of GARCH models. Also, a student t-distribution that can describe skewness and heavy tails has been employed for all GARCH-type models in this study since financial markets typically have fat tails.

2.4. Deep learning models

2.4.1. Long Short-Term Memory neural network (LSTM)

A recurrent neural network is one that receives input sequences and learns sequential patterns through internal loops, in which, in order to learn the weights, the back-propagation is used. Additionally, the chain rule's calculated slope needs to be propagated. The slope gets incredibly little (or extremely huge) as the data are backpropagated into the activation function, such as the sigmoid and tanh functions, and runs into the issue of disappearing (or exploding gradients). Backpropagation is susceptible to remote dependencies. To prevent these issues, LSTM models were created. Hochreiter and Schmidhuber (1997) proposed LSTM, which introduces the use of memory cells and gates to store data over extended periods of time or to delete irrelevant data.

$$Input\ gate: \widetilde{c}^{<t>} = tanh(W_c[a^{t-1}, x^{<t>}] + b_c) \tag{9}$$

$$Updata\ gate: \Gamma_u = \sigma(W_u[a^{t-1}, x^{<t>}] + b_u) \tag{10}$$

$$Forget\ gate: \Gamma_f = \sigma\left(W_f\left[a^{t-1}, x^{<t>}\right] + b_f\right) \tag{11}$$

$$Output\ gate\ 1: \Gamma_o = \sigma\left(W_o\left[a^{t-1}, x^{<t>}\right] + b_o\right) \tag{12}$$

$$Output\ gate\ 2: c^{<t>} = \Gamma_u \cdot \widetilde{c}^{<t>} + \Gamma_f \cdot c^{<t-1>} \tag{13}$$

$$Output\ gate\ 3: a^{<t>} = \Gamma_o \cdot tanh\ c^{<t>} \tag{14}$$

In these equations, $c^{<t>}$ *is memory cell*, $a^{<t>}$ *is activation value*, W is weight matrixes, b is a bias term, σ *(·)* is a sigmoid function, and symbol · denotes element-wise multiplication. Also, the workflow of LSTM plot (see Fig. 1) is displayed below.

First, the Γ_u of forget gate produces the weighted sum of a^{t-1},$x^{<t>}$, and bias as a value from 0 to 1 through the sigmoid function. A value of one passed through the gate means that "all input information passes through the gate", and a value of zero means that "no input information is passed". Thus, as shown in output gate 2, the forget gate controls the amount of information in the past cell state $c^{<t-1>}$ in updating the cell state at time t. The update gate Γ_u determines how much new information is stored in the cell state $c^{<t>}$. The input gate calculates the new information at time t and its output through the tanh function has a value range from −1 to 1. As noted in output gate 2, past cell state information and new information, which are controlled by the forget and input gates, are calculated as variable $c^{<t>}$ of time t. Finally, output value $a^{<t>}$ is determined by passing through the output gate o t input gate and filtering at $c^{<t>}$, while $c^{<t>}$ passes into the tanh function so that the value is between −1 and 1. The selected values are converted to output by multiplying them by Γ_o. This process updates the cell state $c^{<t-1>}$, necessary information is separated from unnecessary information, and output becomes $a^{<t>}$ as noted in output gate 3. The backpropagation via time algorithm is used to learn the LSTM model that is made up of these memory blocks.

2.4.2. Convolutional Neural Network (CNN)

The purpose of the convolutional and pooling layers is to filter the input data and extract relevant information that will be utilized as an input mainly on a fully connected network layer (Rawat & Wang, 2017). To be more precise, the convolutional layers perform a convolution operation between the raw input data and convolution kernels to generate new feature values. Since this method was designed to extract features from picture datasets, the input data is required in structured matrix form (Livieris et al., 2020). When compared to the input matrix, the convolution kernel (filter) is thought of as a narrow window that organizes coefficient values into a matrix. This window ''slides'' across the input matrix, performing a convolution operation on each subregion (patch) that it ''meets''. A convolved matrix—the end result of all these operations—represents a feature value determined by the coefficient values and the filter's applied dimension size. Different convolution kernels can be used to the input data to produce various convolved features, which tend to be more valuable than the original features of the input data and improve the performance of the model.

A nonlinear activation function (such as a rectified linear unit) and a pooling layer are typically placed after the convolutional layers. A pooling layer is a subsampling approach that creates a lower-dimension matrix by extracting specific values from the convolved features. The pooling layer uses a small sliding window in a manner similar to how operations on the convolutional layer are carried out. It takes as input the values of each patch of the convolved features and outputs one new value that is determined by an operation that the pooling layer is defined to carry out.

Briefly, while LSTM layers are used to effectively capture sequence patterns as well as long and short-term dependencies, convolutional layers are used to filter out the noise in complicated time-series data and extract new important features. Overall, when these two layers are assembled together, the prediction performance will be improved tremendously for sequential data. Fig. 2 depicts the architecture of CNN-LSTM.

2.4.3. Hybrid models

Previous research has explored the combination of statistical models and neural networks to enhance the accuracy of volatility forecasting (Kakade et al., 2022; Jiao et al., 2022; Kim & Won, 2018). These studies introduced hybrid models that exhibited improved

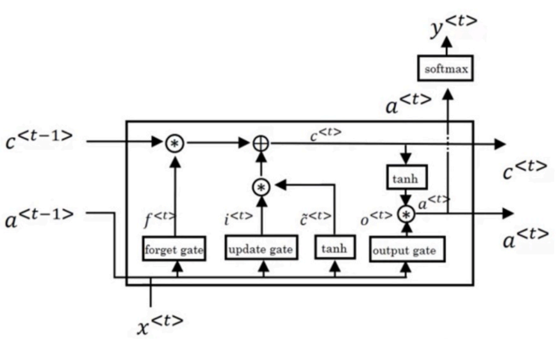

Fig. 1. LSTM structure.

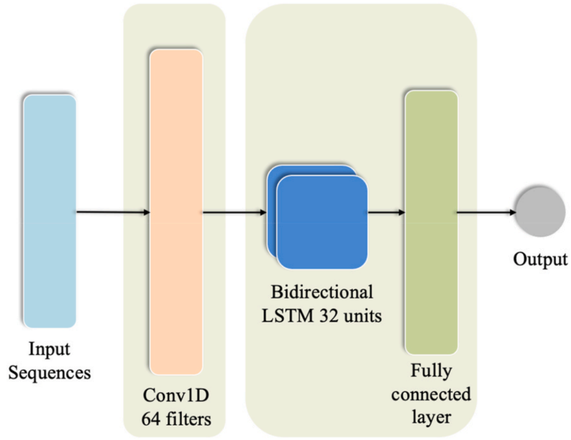

Fig. 2. CNN-LSTM structure.

volatility prediction. Additionally, research by Kakade et al. (2022), Lu et al. (2022), Xu et al. (2022) highlighted the benefits of combining GARCH-type models with LSTM for VaR forecasting. This study follows a similar approach for VaR forecasting construction, and for in-depth technical details, one can refer to Kakade et al. (2022).

The historical volatility information is derived from predicted volatility using GARCH, IGARCH, EGARCH, and GJRGARCH models, each capturing distinct economic characteristics. As explained earlier, each GARCH-type model, including GARCH and IGARCH for volatility clustering and leptokurtosis, and EGARCH and GJRGARCH for leverage effect modeling, captures distinct economic characteristics. The proposed hybrid model, integrating various GARCH-type models with LSTM-based models, aims to enhance VaR forecasting effectiveness by providing comprehensive information for VaR estimation. In this study, for each statistical model, qualitative and quantitative information will be used as independent variables in the construction of LSTM and CNN-LSTM models, accordingly.

The LSTM and CNN-LSTM models are constructed by using Keras and Tensorflow library. The proposed model adopts a bidirectional LSTM framework, featuring layers with 64 and 32 units, employing the ReLU activation function. It incorporates a dropout rate of 0.2, utilizes the Adam optimizer, and minimizes the mean squared error (MSE) loss function. The training process spans 60 epochs with a batch size of 30, includes a 20 % validation split, and enables data shuffling. To optimize the CNN, this study employs widely used optimization algorithms, specifically the Adam optimizer. The architecture combines CNN and LSTM layers to leverage their respective strengths. It starts with TimeDistributed Conv1D layers, configured with 64 filters, a kernel size of 3, and ReLU activation. These are followed by MaxPooling1D layers with a pool size of 2 and Flatten layers to prepare the data for subsequent processing. The LSTM component incorporates bidirectional LSTM layers, each containing 64 units, with dropout rates of 0.5 and 0.2 for regularization to prevent overfitting. A final Dense layer is tailored to match the shape of the target variable, train Y. The model is compiled using the Adam optimizer and a MSE loss function.

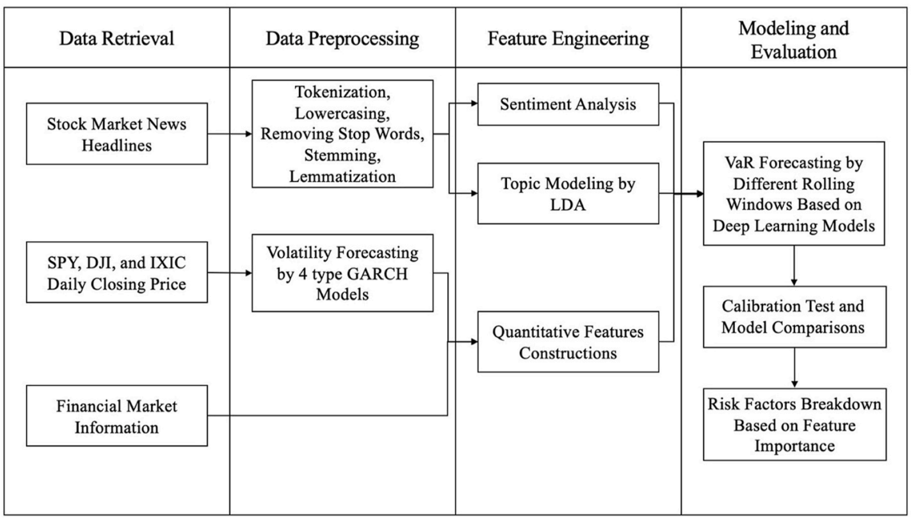

Fig. 3. Research design of this paper.

3. Empirical study

3.1. Research design of this paper

The entire empirical process, depicted in Fig. 3, comprises four segments: data retrieval, data preprocessing, feature engineering, and modeling and evaluation. The initial and second segments involve the preprocessing of textual and non-textual information associated with the three indices, respectively. This results in structured data suitable for text mining and subsequent application in the deep learning model. The third segment involves qualitative and quantitative features construction for the preparation of deep learning models to forecast the three indices VaR. The fourth segment involves forecasting VaR applied the LSTM and CNN-LSTM model implemented two 7-day and 20-day rolling windows implemented, comparing the model results with those from traditional econometric. The 7-day and 20-day rolling windows are chosen because they represent short and long horizon maturity lags, respectively, which have been validated to have different impacts on VaR forecasting (Bouteska et al., 2023). Additionally, the influence of textual data on VaR forecasting is examined, and the significant risk factors are analyzed in detail.

3.2. Data description

Headline news sourced from www.investing.com/news/stock-market-news provided textual information, yielding a dataset of 216,550 news headlines spanning from February 21, 2012, to February 23, 2023. Numerical data, encompassing the opening, high, low, trading volume, and closing prices, was extracted from www.investing.com for constructing other independent variables like AR, BR, and historical volatility over the same period. To facilitate model fitting and evaluation, the dataset was divided into in-sample data (February 21, 2012, to August 31, 2019) for parameter estimation and post-sample data (September 1, 2019, to February 23, 2023) for performance assessment. To enhance model fitting and prediction, we address missing values in the comprehensive dataset and apply min-max scaling normalization after constructing the data from both textual information and numerical data.

Fig. 4 shows the trends of the three indices, revealing that all index prices have experienced dramatic variations since the pandemic outbreak. Remarkably, the DJI shows greater fluctuations than the other indices. As inferred from Table 1, the standard deviation of the out-of-sample period is much larger than that of the in-sample period because the out-of-sample period includes the pandemic. The Jarque-Bera test is rejected for all indices in both periods, as shown in Table 1, indicating that all log returns did not follow a normal distribution in either period. Additionally, the null hypothesis of the ADF test is similarly disproved for all indices, showing that there is no unit root in the log return series for either period.

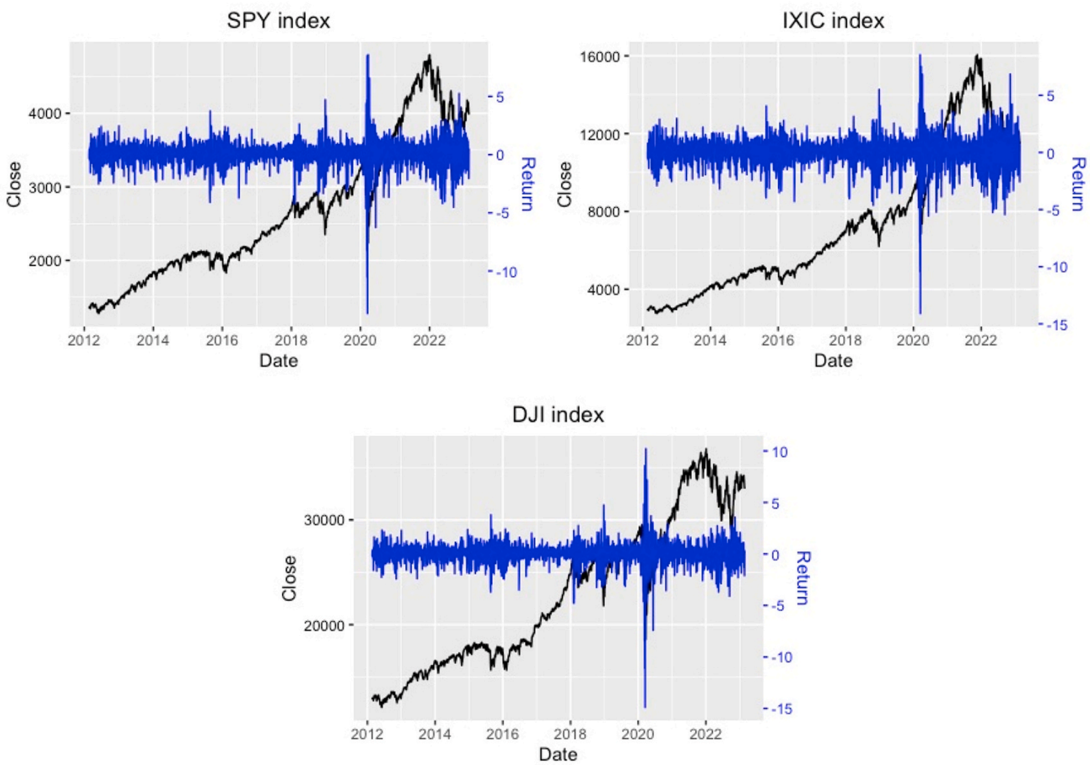

Fig. 4. The daily closing price and returns of three indices *(United States SPY, United States DJI, United States IXIC) from February 21, 2012, until February 23, 2023.* Source: *Yahoo Finance.*

3.3. Grouping features based on LDA

We analyze the diverse effects of news topics through the examination of sub-grouped text features using LDA topics. However, an essential parameter, the number of topics, must be defined before applying LDA. The optimal number of topics relies on dataset characteristics and size. For instance, a larger and diverse dataset may warrant more topics, but for a collection focused on a specific subject, additional similar articles might not introduce more topics. This study employs the Kullback–Leibler (KL) divergence approach suggested by Lin et al., 2010 to ascertain the appropriate number of topics. The selection of the number of topics is based on the highest KL-divergence among topics. Fig. 5 illustrates the Topic KL-divergence for topic numbers ranging from 2 to 20, with the maximum divergence occurring at a topic number of 7.

The LDA model yields seven identified topics, each with its respective distributions of topics, topic words, and word distributions. Table 2 showcases the top 10 words with the highest weights in each topic, providing insights into factors influencing VaR forecasting. Analyzing these words allows the identification of patterns and topic conclusions. The LDA model categorizes news headlines into topics primarily associated with stock market trends and various impact factors. In the subsequent section, these topics, treated as independent variables, are assessed for their potential to enhance VaR prediction performance and to understand their influence on VaR forecasting. Table 3 details the qualitative and quantitative indicators utilized in model construction.

3.4. Evaluation methods

For VaR model backtesting, numerous studies have put forth a variety of loss functions and evaluation techniques. Nevertheless, a number of important characteristics of various models are consistent with a number of evaluation strategies (González-Rivera et al., 2004). As a result, one loss function is utilized in this empirical study to compare the performance of several models, along with the two measurements that are most frequently used to measure VaR accuracy. In our empirical study, we looked at the day-ahead estimation of the VaR for the following usual risk levels: 1 % and 5 %. To evaluate the VaR accuracy in the post-sample period, we carried out a detailed back-testing exercise based on an unconditional coverage test, a conditional coverage test, and a loss function. The risk measurement model is brought to the test using the likelihood ratio, with the Kupiec test provided as one of the major unconditional coverage tests to investigate and assess the model's capability to cover risk (Kupiec, 1995). This is how the corresponding equation is expressed:

$$2\,ln\left[\left(1-\frac{N}{T}\right)^{T-N}\left(\frac{N}{T}\right)^{N}\right]-2\ln\left[(1-p)^{T-N}p^{N}\right]\sim\chi_1^2$$

The observed number of exceptions in the sample is N. T stands for the overall sample size, while p for the anticipated frequency of exceptions. This test has one degree of freedom and is under χ_1^2-distributed. A model can be rejected for both over-high and over-low failure when this test is applied, however, its power could be poor stated by Kupiec (1995). Therefore, we turn to another complementary calibration test.

To evaluate conditional coverage, we used the Engle and Manganelli (2004) dynamic quantile test, which is a joint test of appropriate unconditional coverage and independence of violations. It indicates whether the Hit_t variable's distribution is i.i.d. Bernoulli with probability, is independent of the hit variable's lags and the conditional quantile forecast and is defined as $Hit_t = I(r_t < q_t(\theta)) - \theta$. According to Engle and Manganelli (2004) studies, we used four delays for the Hit variable in the regression test.

In order to better compare the various models' prediction performance, a VaR-based loss function (González-Rivera et al., 2004) is utilized and expressed:

$$Q \equiv P^{-1}\sum_{t=R}^{T}\left(\alpha-d_{t+1}^{\alpha}\right)\left(y_{t+1}-VaR_{t+1}^{\alpha}\right)$$

Where $d_{t+1}^{\alpha}\equiv 1\left(y_{t+1}< VaR_{t+1}^{\alpha}\right)$. The observations for which $y - VaR^{\alpha} < 0$are penalized more severely with weight (1 - α), making this an asymmetric loss function. Goodness of fit is better when Q is smaller.

Table 1
Descriptive statistics of three indices.

	Mean	Std.dev	Skewness	Kurtosis	ADF	Jarque-Bera
In-sample						
SPY	0.0384	0.8158	−0.4638	3.2818	−12.5870	898.62
IXIC	0.0493	0.9764	−0.4797	2.7390	−12.7500	650.91
DJI	0.0359	0.8043	−0.4728	3.5584	−12.2800	1047.30
Out-of-sample						
SPY	0.0221	1.5113	−1.1085	13.6179	−8.7896	7367.40
IXIC	0.0246	1.7604	−0.8433	7.3540	−9.1230	2204.70
DJI	0.0123	1.4927	−1.3890	20.0534	−8.9173	15858.00

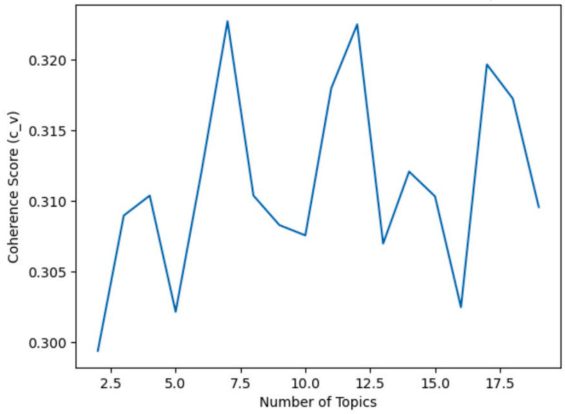

Fig. 5. Topic KL-Divergence of topic numbers, ranging from 2 to 20.

Table 2
Topic keywords and explanation.

Topics	Explanation	Keywords
Market Movements	Stock and trade movements impacting close values. Higher and lower trends with a focus on U.S. markets.	stock, trade, close, higher, us, share, lower, say, billion, fall
Market Trends and China Impact	Stock and trade trends with a focus on higher values. Impact on shares and trends influenced by China.	stock, trade, close, higher, share, lower, us, say, china, beat
Dow, European, and Data Impact	Stock trade and close trends with a focus on Dow and Dax. Lower trends influenced by European market and data.	stock, us, dow, lower, trade, european, close, dax, jone, datum
Earning, Impact	Stock trade and close trends with impacts on earnings. Higher and lower trends influencing shares and revenue.	stock, trade, close, lower, higher, share, us, beat, earn, revenue
Higher and Lower Movements	Stock trade and close trends with higher and lower values. Impact on shares, buying trends, and earnings.	stock, close, trade, higher, lower, share, us, buy, earn, say
Earnings Mix	Stock and trade trends with a mix of earnings. Influences on revenue, earnings, and a mix of trade activities.	stock, beat, revenue, us, earn, trade, close, mix, dow, say
Higher, Lower, and Banking Trends	Stock trade and close trends with higher and lower values. U.S. market trends, impacts on shares, and trends in banking.	stock, trade, close, higher, lower, us, share, china, 3, bank

Table 3
List of 19 independent variables.

Indicator	Explanation
AR	The market's popularity
BR	The willingness of the market to buy and sell
Price movement	Intraday price fluctuations
Risk factor classification	The risk information of news headlines
Emotional polarity score	The daily sentiment of news headlines.
Subjectivity score	The subjectivity of news headlines
Historical volatility information	Lagged predicted volatility from GARCH-type models
InVol	Log trading volume
IndVol	Indicator for high/low volume
Neg news headline	The negative news headlines ratio
Neutral news headline	The neutral news headlines ratio
Positive news headline	The positive news headlines ratio
Topic 1	Market movements
Topic 2	Market trends and China impact
Topic 3	Dow, European, and data impact
Topic 4	Earnings impact
Topic 5	Higher and lower movements
Topic 6	Earnings mix
Topic 7	Higher, lower, and banking trends

4. Results and discussion

Table 4 presents the results of both unconditional and conditional tests for out-of-sample data, comparing benchmark methods with hybrid models using 7-day rolling windows. The table also provides a detailed analysis of the significance of textual indicators by comparing VaR forecasting outcomes with and without textual components in our models. Among conventional models, all benchmarks underestimate risk at a 95 % confidence level according to Kupiec's test and face many rejections in the DQ test. Most conventional models perform well at a 99 % confidence level, with the exception of the EGARCH model, which is almost always rejected by both calibration tests. It is evident that combining traditional methods with deep learning models enhances overall performance. A key insight from Table 4 is that excluding text variables from hybrid LSTM and CNN-LSTM models leads to more rejections in both unconditional and conditional tests at both confidence levels, indicating a diminished predictive performance in the absence of textual information.

Moving to Table 5, traditional methods combined with LSTM and CNN-LSTM exhibit substantial improvement via passing all calibration tests on previous basic, achieving the lower loss values compared to conventional models. Besides, Table 5 becomes an evident that the inclusion of textual features leads to an apparent reduction in average loss values within hybrid LSTM and CNN-LSTM models, particularly remarkable at a 99 % confidence level. Crucially, the CNN-LSTM models, harnessing the noise-filtering capability of CNN, exhibit smaller loss values, pointing toward an improved prediction when incorporating textual variables. While LSTM models are widely accepted for time series forecasting, incorporating additional convolutional layers significantly enhances forecasting performance, aligning with findings in the gold price forecasting study by Livieris et al. (2020). Additionally, Table 5 highlights that EGARCH-CNN-LSTM perform best at both 95 % and 99 % confidence levels, owing to their lowest loss values.

Bouteska et al. (2023) found a statistically significant correlation between online investor sentiment and the VaR of individual stocks, particularly evident with a shorter trading horizon (7-day horizon). To investigate the impact of textual data on VaR forecasting in different horizons, we employ 7-day and 20-day rolling windows when the DL models were implemented. Table 6 provides calibration tests for deep learning models across these rolling windows and compares the loss value reduction with benchmark models. As per the results of the calibration tests, it is evident that the hybrid LSTM and CNN-LSTM models with 20-day rolling windows outperform their counterparts with 7-day rolling windows. Additionally, the hybrid models, specifically the LSTM and CNN-LSTM models, exhibit smaller loss values with the 20-day rolling windows, suggesting improved VaR estimation. This observation implies that longer rolling windows contribute to better VaR estimation, possibly due to the ability of convolutional layers to extract valuable textual information and the effectiveness of LSTM networks in capturing both short-term and long-term dependencies. The longer rolling windows likely provide more information, leveraging these mechanisms for enhanced performance.

To assess the impact of textual data on VaR estimation across different rolling windows, the feature importance algorithm identifies a subset of features that can be utilized for achieving the most accurate forecasting (Genuer et al., 2010). Consequently, a high-ranking feature implies a significant influence on VaR forecasting. In our research, the feature importance in deep learning models is determined by the average rank across features in each hybrid model. The varying degrees of influence based on feature ranks are presented in Table 7. Fig. 6 compares the differences in the features importance of textual data for three indexes, using 7-day and 20-day rolling

Table 4
Results of VaR models and hybrid models regarding the Kupice's test and DQ test.

	95 % Confidence Level		99 % Confidence Level	
	Kupice's Test	DQ Test	Kupice's Test	DQ Test
Conventional Models				
GARCH	3	3	1	0
IGARCH	3	1	1	0
GJRGARCH	3	2	1	1
EGARCH	3	3	2	3
Without Textual Data (7 Days Rolling Windows)				
GARCH-LSTM	0	1	0	0
IGARCH-LSTM	0	0	0	0
GJRGARCH-LSTM	0	0	0	0
EGARCH-LSTM	0	0	0	0
GARCH-CNN-LSTM	0	0	0	0
IGARCH-CNN-LSTM	0	0	1	0
GJRGARCH-CNN-LSTM	0	0	0	0
EGARCH-CNN-LSTM	0	0	0	0
With Textual Data (7 Days Rolling Windows)				
GARCH-LSTM	0	0	0	0
IGARCH-LSTM	0	0	0	0
GJRGARCH-LSTM	0	0	0	0
EGARCH-LSTM	0	0	0	0
GARCH-CNN-LSTM	0	0	0	0
IGARCH-CNN-LSTM	0	0	0	0
GJRGARCH-CNN-LSTM	0	0	0	0
EGARCH-CNN-LSTM	0	0	0	0

Note: Smaller value is preferred.

Table 5
Comparison of hybrid model performance in loss value reduction with and without the inclusion of textual Variables.

Examine variable influence	Color	Feature rank
Strong Impact		1~7
Medium Impact		8~13
Weak Impact		14~19

Table 6
Calibration tests and loss value reduction in deep learning models: A comparison across 7 and 20 rolling windows with benchmark models".

	7-day rolling windows			20-day rolling windows		
	Kupiec's Test	DQ test's P-value	Average Loss Value Reduction	Kupiec's Test	DQ test's P-value	Average Loss Value Reduction
95 % confidence level:						
LSTM Hybrid Model	0	0	−7.19 %	0	0	−11.87 %
CNN-LSTM Hybrid Model	0	0	−10.15 %	0	0	−15.65 %
Total:	0	0	−8.67 %	0	0	−13.76 %
99 % confidence level:						
LSTM Hybrid Model	0	0	−9.31 %	0	0	−15.60 %
CNN-LSTM Hybrid Model	0	0	−12.68 %	0	0	−17.20 %
Total:	0	0	−10.99 %	0	0	−16.40 %

Note: Smaller value is preferred.

windows. Particularly, a more pronounced influence of strong impact textual data on hybrid models is notably observed in the 20-day rolling windows. This indicates that providing robust impact textual data is crucial for accurate VaR estimation, emphasizing the importance of feature selection in improving forecasting accuracy across different time horizons. Additionally, Table 7 reveals an interesting observation: the CNN-LSTM hybrid models show fewer topics with weaker impact for both rolling windows. This highlights that for this approach, the key to generating accurate VaR forecasts is not in increasing the quantity of input variables but in the model's ability to identify the independent variables with strong impact.

For a thorough analysis of the consistent and varying impacts of textual data across different rolling windows, Table 8 examines the influence of textual data features. Topic 6, labeled "Earnings Mix," and Topic 1, labeled "Market Moving," consistently play a crucial role in VaR forecasting. Specifically, the integration of stock and trade trends with earnings-related information, such as news about revenue, earnings, and a mix of trade activities, significantly influences VaR forecasting. Additionally, movements in stock and trade impacting closing values, particularly higher and lower trends in U.S. markets, contribute significantly to VaR forecasting, regardless of the rolling window duration. Emotional polarity scores also emerge as a significant factor, especially in longer 20-day rolling windows, suggesting that an extended time horizon is necessary for these scores to influence VaR estimation, possibly due to the long-memory characteristics of news headline sentiments. Furthermore, Topic 2, labeled "Market Trends and China Impact," strongly influences VaR forecasting through CNN-LSTM, particularly for the SPY and IXIC indexes. This underscores the substantial role of Topic 2 in most models, with some exceptions, highlighting its potential significance in VaR estimation.

For a more comprehensive understanding of the trends associated with each topic, we analyzed monthly tendencies by averaging the relevance of each topic within daily news headlines. Figs. 7 and 8 illustrate the trends for the most and least significant topics, respectively. In Fig. 7, the highly significant topic "Earnings Mix" (Topic 6) has seen a decline in attention; however, its prominence alone is insufficient for it to play a critical role. In contrast, another significant topic, "Market Movements" (Topic 1), shows a minor

Table 7
Textual Data Impact on VaR Estimation: Feature Importance Insights.

Examine variable influence	Color	Feature rank
Strong Impact		1~7
Medium Impact		8~13
Weak Impact		14~19

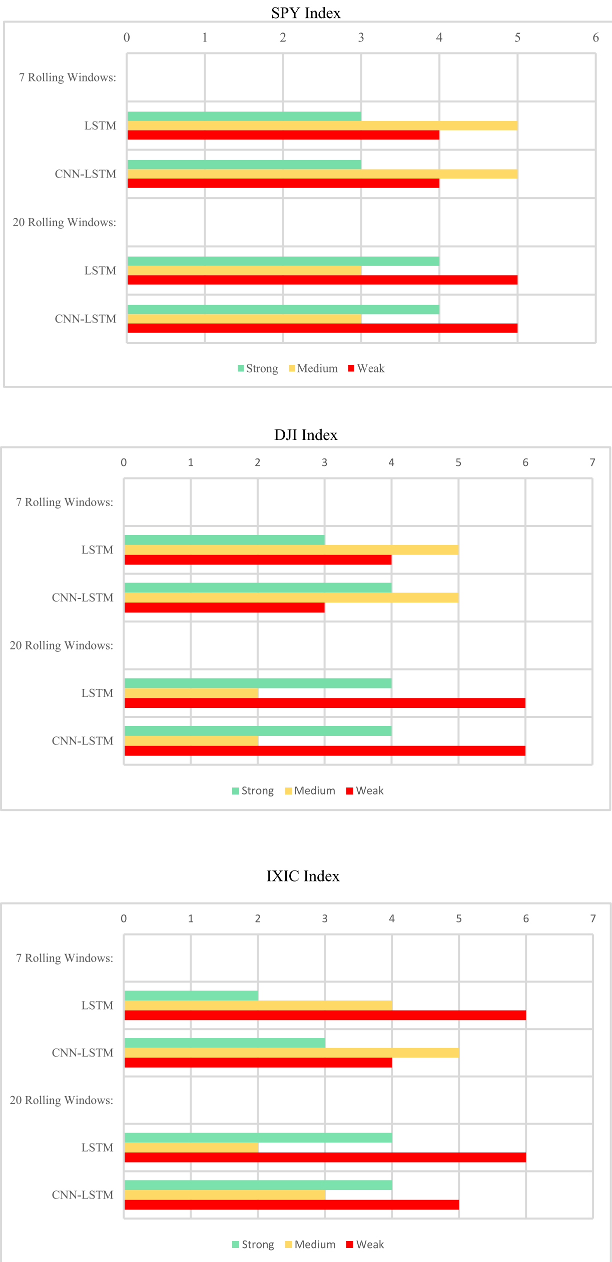

Fig. 6. The quantity of textual data feature importances across different rolling windows for three stock indices.

upward trend.

Regarding the less important topics, all have garnered considerable attention and experienced substantial increases from 2012 to 2023, as shown in Fig. 8. While these topics may occasionally provide valuable information to investors, their contribution may not significantly enhance our understanding and accurate measurement of risk. In fact, the percentage of daily news related to the important topics “Earnings Mix” and “Market Movements” is lower than 92 %, whereas daily news reports on the less important topics exceed 96 %. Therefore, it is advisable to focus on a select few important topics when constructing a VaR prediction model and encourage stock market participants to pay closer attention to these topics to improve risk assessment.

5. Conclusion

This study introduces a novel VaR forecasting model by combining text mining methods and deep learning techniques. While the concept of hybrid models incorporating text mining has been explored in various fields (Jiao et al., 2022; Kakade et al., 2022; Kim &

Table 8
Influence of Textual Data Features Across Varied Rolling Windows.

SPY	7 Rolling Windows:		20 Rolling Windows:	
Textual Features:	LSTM	CNN-LSTM	LSTM	CNN-LSTM
1. Emotional Polarity Score	Yellow	Red	Green	Green
2. Subjectivity Score	Yellow	Yellow	Red	Yellow
3. Negative News Headline	Red	Red	Red	Yellow
4. Neutral News Headline	Red	Red	Red	Red
5. Positive News Headline	Yellow	Red	Red	Red
6. Market Movements	Green	Green	Green	Green
7. Market Trends and China Impact	Red	Yellow	Yellow	Green
8. Dow, European, And Data Impact	Red	Yellow	Yellow	Yellow
9. Earnings Impact	Green	Yellow	Green	Red
10. Higher And Lower Movements	Yellow	Yellow	Yellow	Red
11. Earnings Mix	Green	Green	Green	Green
12. Higher, Lower, And Banking Trends	Yellow	Green	Red	Red
DJI	LSTM	CNN-LSTM	LSTM	CNN-LSTM
1. Emotional Polarity Score	Yellow	Green	Green	Green
2. Subjectivity Score	Yellow	Yellow	Red	Yellow
3. Negative News Headline	Red	Red	Red	Red
4. Neutral News Headline	Red	Red	Red	Red
5. Positive News Headline	Yellow	Yellow	Red	Yellow
6. Market Movements	Green	Green	Green	Green
7. Market Trends and China Impact	Yellow	Yellow	Yellow	Red
8. Dow, European, And Data Impact	Yellow	Yellow	Green	Green
9. Earnings Impact	Green	Green	Yellow	Red
10. Higher And Lower Movements	Red	Red	Red	Red
11. Earnings Mix	Green	Green	Green	Green
12. Higher, Lower, And Banking Trends	Red	Yellow	Red	Red
IXIC	LSTM	CNN-LSTM	LSTM	CNN-LSTM
1. Emotional Polarity Score	Red	Yellow	Green	Green
2. Subjectivity Score	Yellow	Yellow	Green	Yellow
3. Negative News Headline	Red	Red	Red	Red
4. Neutral News Headline	Red	Red	Red	Red
5. Positive News Headline	Red	Red	Yellow	Yellow
6. Market Movements	Red	Yellow	Red	Green
7. Market Trends and China Impact	Yellow	Green	Yellow	Green
8. Dow, European, And Data Impact	Yellow	Red	Red	Red
9. Earnings Impact	Green	Green	Green	Yellow
10. Higher And Lower Movements	Red	Yellow	Red	Red
11. Earnings Mix	Green	Green	Green	Green
12. Higher, Lower, And Banking Trends	Yellow	Yellow	Red	Red

Won, 2018; Livieris et al., 2020), it has received limited attention in VaR forecasting. Through empirical analysis, we compared GARCH-type models with hybrid models that integrate LSTM and CNN-LSTM models. Our findings underscore the significance of integration of benchmark models and DL models, which act as powerful enhancers for improving VaR estimation. Importantly, CNN-LSTM models, leveraging CNN's noise-filtering capabilities, exhibit lower loss values, indicating substantially improved predictive capabilities. Our research also reveals that longer rolling windows contribute to better VaR estimation, likely attributed to the convolutional layers' capacity to extract valuable textual insights and the effectiveness of LSTM networks in capturing both short-term and long-term dependencies. Moreover, we confirmed that including textual variables as independent factors significantly enhances VaR forecasting performance. Furthermore, our study unveils two notable findings related to textual data. Firstly, emotional polarity scores gain prominence in longer rolling windows, suggesting that they require an extended horizon to exert a notable impact on VaR estimation, likely due to the long-memory characteristics of news headline sentiments. Secondly, it is advisable to focus on a select few

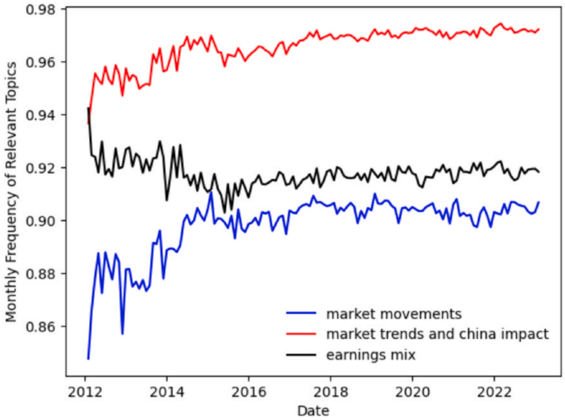

Fig. 7. Monthly tendency of important topics.

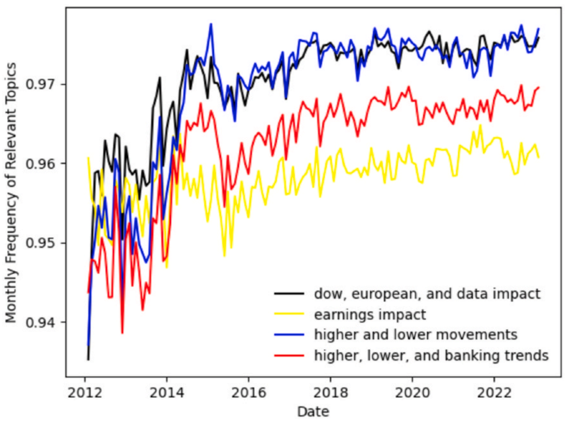

Fig. 8. Monthly tendency of less important topics.

key news topics, such as “Market Movements,” “Earning Mix,” and “Market Trends and China Impact,” when constructing a VaR prediction model. This approach encourages stock market participants to prioritize these topics for more precise risk measurement and preparation. These results have important implications for risk practitioners: by integrating NLP-derived news signals into standard VaR frameworks, risk managers can achieve more timely and accurate capital allocation and stress-testing. Theoretically, our work contributes to the growing literature on hybrid deep-learning risk models, demonstrating how textual features can be systematically incorporated to advance both micro prudential risk forecasting and the development of best-practice regulatory standards.

One limitation of this study is its exclusive focus on VaR as the risk metric. While VaR remains widely adopted due to its simplicity, interpretability, and well-established backtesting procedures, future research could extend this NLP-based framework to more complex systemic risk measures such as Expected Shortfall. Beyond this, future work could explore the integration of high-frequency data, leveraging its granularity and the capacity of deep learning models to handle large datasets effectively. Additionally, delving into deep learning models combined with unconventional textual sources like market news comments presents an intriguing avenue. Another prospect lies in adapting the proposed Structural Topic Model (STM) that has the capability to integrate metadata alongside topic modeling. This feature enables it to consider variables such as time, authorship, or any other metadata that can impact how topics are distributed within documents. This is valuable for examining how topics evolve over time or in diverse settings.

Author statement

Cao Yangfan: Conceptualization, Methodology, Software, Writing- Original draft preparation. Choo Wei Chong: Visualization, Investigation. Bolaji Tunde Matemilola: Supervision: Writing- Reviewing and Editing.

Funding information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Angelidis, T., Benos, A., & Degiannakis, S. (2004). The use of GARCH models in VaR estimation. *Statistical Methodology, 1*, 105–128.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometricseconometrics, 31*, 307–327.

Bouteska, A., Cardillo, G., & Harasheh, M. (2023). Is it all about noise? Investor sentiment and risk nexus: Evidence from China. *Finance Research Letters, 57*, Article 104197.

Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the multicollinearity problem and its machine learning approach: A review. *Mathematics, 10*(8), 1283.

Choi, J. E., & Shin, D. W. (2022). Parallel architecture of CNN-bidirectional LSTMs for implied volatility forecast. *Journal of Forecasting, 41*(6), 1087–1098.

Deveikyte, J., Geman, H., Piccari, C., & Provetti, A. (2022). A sentiment analysis approach to the prediction of market volatility. *Frontiers in Artificial Intelligence, 5*, Article 836809.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society, 50*, 987–1007.

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

Engle, R. F., & Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance, 1*, 237.

Figà-Talamanca, G., & Patacca, M. (2024). An explorative analysis of sentiment impact on S&P 500 components returns, volatility and downside risk. *Annals of Operations Research, 342*(3), 2095–2117.

Fuertes, A. M., Kalotychou, E., & Todorovic, N. (2015). Daily volume, intraday and overnight returns for volatility prediction: Profitability or accuracy? *Review of Quantitative Finance and Accounting, 45*, 251–278.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*(14), 2225–2236.

Giacometti, R., Bertocchi, M., Rachev, S. T., & Fabozzi, F. J. (2007). Stable distributions in the black–litterman approach to asset allocation. *Quantitative Finance, 7*(4), 423–433.

González-Rivera, G., Lee, T. H., & Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting, 20*(4), 629–645.

Gurrib, I., & Kamalov, F. (2022). Predicting bitcoin price movements using sentiment analysis: A machine learning approach. *Studies in Economics and Finance, 39*(3), 347–364.

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

Jiao, X., Song, Y., Kong, Y., & Tang, X. (2022). Volatility forecasting for crude oil based on text information and deep learning PSO-LSTM model. *Journal of Forecasting, 41*(5), 933–944.

Jiao, P., Veiga, A., & Walther, A. (2020). Social media, news media and the stock market. *Journal of Economic Behavior & Organization, 176*, 63–90.

Kakade, K., Jain, I., & Mishra, A. K. (2022). Value-at-Risk forecasting: A hybrid ensemble learning GARCH-LSTM based approach. *Resources Policy, 78*, Article 102903.

Kakade, K., Mishra, A. K., Ghate, K., & Gupta, S. (2022). Forecasting commodity market returns volatility: A hybrid ensemble learning GARCH-LSTM based approach. *Intelligent Systems in Accounting, Finance and Management, 29*(2), 103–117.

Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications, 103*, 25–37.

Koukaras, P., Nousi, C., & Tjortjis, C. (2022). Stock market prediction using microblogging sentiment analysis and machine learning. *Tele.com, 3*(2), 358–378.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. In *Division of research and statistics*. Division of Monetary Affairs. Federal Reserve Board.

Li, X., Shang, W., & Wang, S. (2019). Text-based crude oil price forecasting: A deep learning approach. *International Journal of Forecasting, 35*(4), 1548–1560.

Lin, S. H., Yeh, Y. M., & Chen, B. (2010). Leveraging kullback–leibler divergence measures and information-rich cues for speech summarization. *IEEE Transactions on Audio Speech and Language Processing, 19*(4), 871–882.

Liu, M., Taylor, J. W., & Choo, W. C. (2020). Further empirical evidence on the forecasting of volatility with smooth transition exponential smoothing. *Economic Modelling, 93*, 651–659.

Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN–LSTM model for gold price time-series forecasting. *Neural Computing & Applications, 32*, 17351–17360.

Lu, C., Teng, Z., Gao, Y., Wu, R., Hossain, M. A., & Fang, Y. (2022). Analysis of early warning of RMB exchange rate fluctuation and value at risk measurement based on deep learning. *Computational Economics, 59*(4), 1501–1524.

Markowitz, H. (1952). Modern portfolio theory. *The Journal of Finance, 7*(11), 77–91.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science, 60*(7), 1772–1791.

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation, 29*(9), 2352–2449.

Slof, D., Frasincar, F., & Matsiiako, V. (2021). A competing risks model based on latent dirichlet allocation for predicting churn reasons. *Decision Support Systems, 146*, Article 113541.

Vidal, A., & Kristjanpoller, W. (2020). Gold volatility prediction using a CNN-LSTM approach. *Expert Systems with Applications, 157*, Article 113481.

Xu, Z., Zeng, Y., Xue, Y., & Yang, S. (2022). Early warning of Chinese Yuan's exchange rate fluctuation and value at risk measure using neural network joint optimization algorithm. *Computational Economics, 60*, 1293–1315.

Zhang, C. X., Li, J., Huang, X. F., Zhang, J. S., & Huang, H. C. (2022). Forecasting stock volatility and value-at-risk based on temporal convolutional networks. *Expert Systems with Applications, 207*, Article 117951.