QuanSIRA: The Quantitative Investment Risk Modeling in Stock Markets with Large Language Models

Anonymous ACL submission

Abstract

Stock market analysis is important for investors to make financial decisions. Stock price prediction is widely investigated in the natural language processing area due to the superiority of large language models. Recent works have developed several datasets for stock price predictions. However, investment risk, considered an essential factor for investors, is rarely discussed in NLP applications, and there are limited datasets for investment risk analysis. In this work, we propose methods to quantify investment risk and introduce the dataset **QuanSIRA**. Using this benchmark, we investigate the applications of large language models in tackling quantitative investment risk analysis. The experimental results show the difficulty of investment risk analysis. The model built on pre-trained large language models obtained F1 scores of 68.07 and 65.01 in the in-stock benchmark and the cross-stock benchmark of investment risk prediction task.

1 Introduction

Stock market analysis is essential for investors to make informed decisions, identify opportunities, and ultimately achieve their financial objectives. Recent works on stock market analysis focus on stock price prediction (Zou et al., 2022; Wang et al., 2024; Wen et al., 2024).[1] Based on these predictions, investors can maximize their benefits and make rational allocations of limited resources (PH and Rishad, 2020).

Large-scale textual information about the stock markets, such as Twitter comments (Swathi et al., 2022), financial news (Khan et al., 2022), and policies (Li et al., 2020), are valuable resources for stock market analysis (Fataliyev et al., 2021). Combined with historical stock prices, these models achieve significant improvements in the task of stock price prediction (Chen et al., 2022). Recently, the widespread success of large language models in various text processing tasks has ushered in a new training paradigm. Recent models based on large language models demonstrate superiority in capturing stock price values (Dong et al., 2020) and the curves of stock movements (Zhao et al., 2022b).

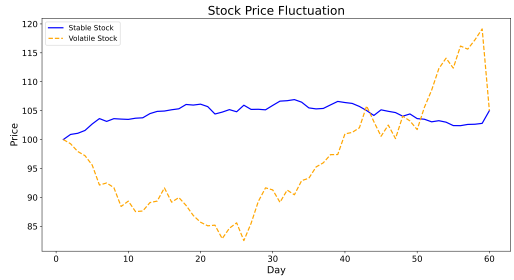


Figure 1: Examples of stock price curves. The dashed curve suffers from higher investment risk compared to the solid curve.

However, accurate predictions of stock prices are not enough to support financial decisions. According to financial analysis (Sonkavde et al., 2023), investment risk should be considered an essential factor for investors when making decisions. Investment risk refers to the uncertainty or probability of losing capital on an investment in stocks. It encompasses various factors that can negatively affect the value of a stock and consequently lead to financial loss for the investor. For example, as shown in Figure 1, although the stocks reach similar prices within the same period, the dashed curve suffers from more risk compared to the solid curve because the prices on the solid curve have a stable increase. In this work, we aim to model investment risk, including the phenomenon of price volatility illustrated in the above example.

Recent works have built several datasets for stock price prediction (Farimani et al., 2021; Sinha

[1] Stock price prediction includes predictions of stock price values and stock price movements.

et al., 2022). However, few works focus on building datasets for stock investment risk analysis due to the difficulty in quantifying investment risk. Aiming to model investment risk that can contribute to financial decision-making, we quantify investment risk according to several factors widely discussed in economics. We focus on short-term investment risk by considering price volatility (Naik and Mohan, 2021), yield deviation (Abbas et al., 2019), systematic volatility (Pelger, 2020), and maximum drawdown (Gupta and Chaudhary, 2023).

Following the release of a comprehensive dataset for stock movement prediction from tweets and historical stock prices (Xu and Cohen, 2018), we introduce the dataset **QuanSIRA** for **Quan**titative **S**tock **I**nvestment **R**isk **A**nalysis.

Based on the proposed dataset, we investigate the application of large language models to tackle stock investment risk analysis by introducing two benchmarks: in-stock investment risk prediction and cross-stock investment risk prediction. The experimental results show that the performance of the investment risk prediction model built on pre-trained language models is marginally acceptable. The task is challenging, and the corresponding models have room for improvement. The contributions of this work are concluded as follows:

- We propose a method to quantify stock investment risk that can contribute to the analysis of stock markets.
- We introduce a dataset with two main benchmarks that can be used to develop novel models for stock investment risk predictions.
- We investigate the applications of various large language models for the quantitative analysis of investment risk. Data and codes are released at `http://anonymous.`

2 Related Work

2.1 Investment Risk Analysis

Investment risk analysis in the stock market is widely discussed as an important aspect of economics. Venturini (2022) investigated the ramifications of climate change on investment risk. Lopez-Lira (2023) analyzed the investment risks inherent in annual disclosures. Dong et al. (2023) examined the impacts of geopolitical, economic, and climate policy risks on energy stocks. Ilbahar et al. (2022) mitigated the impact of expert bias on renewable energy investment risks. Wen et al. (2022) revealed the sensitivity of risk contagion in the markets of oil, stocks, and commodities. Zakhidov (2024) elucidated the roles of various economic indicators in risk assessment.

However, these works focus on high-level discussions of investment risk without quantifying it. In this work, we propose data-driven methods to analyze and predict fine-grained and coarse-grained investment risk.

2.2 Stock Prediction with Pre-trained Language Models

Recent works utilizing various pre-trained language models demonstrate significant improvements in stock prediction compared to previous approaches. Several works focus on building domain-specific financial language models, such as FinBERT (Araci, 2019; Yang et al., 2020; Liu et al., 2021), FLANG (Shah et al., 2022), BloombergGPT (Wu et al., 2023), FinMA (Xie et al., 2023), InvestLM (Yang et al., 2023), and FinGPT (Wang et al., 2023). Due to the effectiveness of sequential modeling, pre-trained language models are used to obtain textual information representations that serve as inputs for stock prediction models. This textual information includes financial news (Dong et al., 2020; Sonkiya et al., 2021), comments from social media (Li et al., 2021; Colasanto et al., 2022), and expert opinions (Zhao et al., 2022a).

All of these works focus on stock price prediction. However, there is limited research on the quantitative modeling and prediction of investment risk, which is the primary focus of this paper.

2.3 Dataset for Stock Prediction

Several datasets have been built for stock prediction in natural language processing. Remy and Ding (2015) collected financial news from Bloomberg and Reuters and demonstrated that financial events are essential for improving stock predictions (Ding et al., 2015). Xu and Cohen (2018) built a comprehensive dataset for predicting stock movements using tweets and historical stock prices.

Previous studies have shown that sentiment analysis can enhance stock prediction. Consequently, sentiment annotations have been included in stock dataset construction efforts (Cortis et al., 2017; Lutz et al., 2018; Farimani et al., 2021; Sinha et al., 2022). To expand these datasets, Dong et al. (2024) collected millions of stock prices and time-aligned financial news records to create a large-scale integration dataset for financial news and stock prices.

However, all of the above datasets include stock price annotations without addressing investment risk. To fill this gap, we propose a method for quantifying investment risk in the stock market and building a dataset with investment risk annotations, which will benefit the natural language processing research community.

3 Preliminary

3.1 Investment Risk

Investment risk pertains to the possibility of experiencing financial losses or receiving returns that fall below expectations due to unpredictable factors. Investment risks in the stock market are categorized into several types, with the primary categories being market risk (Sharpe, 1964) and specific risk (Tirole, 2010).

Market risk includes different types of risks that impact the entire market. These include equity risk (Sharpe, 1964), which causes fluctuations in stock prices due to market movements; interest rate risk (Fabozzi and Fabozzi, 2021), associated with variations in interest rates that affect the valuation of both fixed-income securities and equity investments; and currency risk (Shapiro and Hanouna, 2019), stemming from changes in exchange rates that impact the stock market by affecting the value of international investments and assets denominated in foreign currencies.

Specific risk encompasses a range of risks that affect individual companies rather than the overall market. Business risk (Tirole, 2010) refers to the potential for financial loss or negative impact on a company's operations due to factors inherent in its business activities. Financial risk (Brealey et al., 2014) refers to the various uncertainties and potential losses a company may face in its financial operations. Operational risk (Griffiths, 2016), another significant component, refers to the possibility of issues arising within a company's day-to-day operations. It involves the risk of loss resulting from inadequate or failed internal processes, people, systems, or from external events.

In this paper, we focus on investment risks in the stock market, aiming to analyze and model these risks through the application of large language models. Our research is concerned with market risk, analyzing investment risks by examining market price information and sentiment conveyed in social media.

3.2 Stock Prediction

Stock prediction, an essential component of financial analysis, involves utilizing relevant stock data and applying various analytical methods to forecast future trends in the stock market. To effectively predict the stock market, a multitude of factors and indicators are considered. These factors and indicators include market price data (Fama, 1970), encompassing stock prices, trading volumes, and various market indices, market sentiment data (Bollen et al., 2011) and events data (Bodie and Kane, 2020) such as geopolitical events, regulatory changes, natural disasters, and other influential factors that significantly impact market stability and investor behavior.

The stock prediction is modeled as a traditional classification task. Given a sequential of historical prices, $X = [x_1, x_2, \ldots]$, where x_i is a price of stocks in i-th time steps, and the textual information, $T = [t_1, t_2, \ldots]$, where t_i is i-th word in the text. Both the sequential X and T are constructed utilizing a uniform time step interval. The objective is maximizing the condition probability $P(x_j \mid X_{<j}, T)$.

4 QuanSIRA

In this section, we present QuanSIRA, a comprehensive benchmark developed to quantitatively evaluate and analyze investment risk.

4.1 Investment Risk Quantification

We systematically examine investment risk from four perspectives: price volatility, yield deviation, systematic volatility, and maximum drawdown.

Price Volatility Price volatility is used to measure the changes in stock returns. By analyzing the volatility of returns, we assess the investment risk of a stock. High volatility is typically indicative of market uncertainty, which can precipitate market trepidation, thereby prompting investors to make erroneous decisions, and in turn exacerbates market instability. Higher return volatility indicates greater fluctuations in stock prices, corresponding to increased investment risk. We quantify the stock price volatility V_j on j-th day as:

$$V_j = \sqrt{\frac{1}{\Delta d - 1} \sum_{i=j+1}^{j+\Delta d} (R_i - \bar{R})^2}, \qquad (1)$$

where R_i is the return rate of a stock on i-th day:

$$R_i = \frac{x_i - x_{i-1}}{x_{i-1}}, \tag{2}$$

where x_i is the closing price of a stock on i-th day, and $\bar{R}$ is the average return rate of a stock in the time window $[j+1, j+\Delta d]$:

$$\bar{R} = \frac{1}{\Delta d} \sum_{i=j+1}^{j+\Delta d} R_i. \tag{3}$$

Systematic Deviation The discrepancy between stock returns and the broader market is measured by systematic deviation. We use the beta coefficient as an indicator to measure the deviation of an individual stock's returns from the returns of the broader market. A higher beta coefficient implies that an individual stock is more sensitive to market signals. Consequently, when there is volatility in the overall market, the fluctuation of the individual stock will exceed that of the market, leading to increased investment risk. A larger deviation indicates a significant divergence from the market trend, implying higher investment risk. We calculate the beta coefficient β_j on j-th day as:

$$\beta_j = \frac{\mathrm{Cov}(\bar{R}, \bar{R}_m)}{\sigma_m^2}, \tag{4}$$

where σ_m is the market yield rate variance that is calculated as:

$$\sigma_m^2 = \frac{1}{\Delta d} \sum_{i=1}^{n} (R_i - \bar{R}_m)^2. \tag{5}$$

Yield Deviation The discrepancy between an actual yield rate and its expected yield rate can be measured by yield deviation. The expected yield rate is calculated using the capital asset pricing model (Fama and French, 1992), while the actual yield is derived from the stock's historical observed performance. When the actual yield of a stock significantly deviates from its expected yield, it indicates that the stock performance has not met market expectations. This discrepancy serves as a risk for the stock. A larger deviation indicates a significant divergence from the expected trend, implying higher investment risk. We quantify the yield deviation D_j on j-th day as:

$$D_j = \left| E_j - \bar{R} \right|, \tag{6}$$

where E_j is the expected rate of returns on the j-th day that can be computed as :

$$E_j = R_f + \beta_{j-4:j} \times (\bar{R}_m - R_f), \tag{7}$$

where R_f is the risk-free rate of return that is computed by the average yield rate of U.S. Treasury bills with a 3-month maturity. [2] $\bar{R}_m$ is the average yield rate of the market based on S&P 500. [3] $\beta_{j-4:j}$ is the beta coefficient which is calculated in time interval $[j-4, j]$.[4]

Maximum Drawdown The discrepancy between stock highest price and lowest price is measured by maximum drawdown. The substantial maximum drawdown indicates that within the observed time frame, the stock price has undergone a precipitous decline, akin to a "cliff drop", which could expose investors to heightened risks of losses. This metric reflects the greatest adverse fluctuation in the stock price over a specific period, embodying the investment risk and potential maximal loss. A larger discrepancy indicates a greater drop in stock price, which corresponds to higher investment risk. We quantify the maximum drawdown M_j on j-th day as:

$$M_j = \frac{\max(X) - \min(X)}{\max(X)}, \tag{8}$$

where X is the sequential the closing prices of a stock, $X = [x_{j+1}, x_{j+2}, \ldots, x_{j+\Delta d}]$.

4.2 Investment Risk Taxonomy

We perform min-max normalization on these four investment risk indicators and we calculate the weighted sum of these normalized indicators to obtain the composite metric for assessing investment risk considering these four risk indicators with equal importance.

$$r_j = \frac{N(V_j) + N(D_j) + N(\beta_j) + N(M_j)}{4}, \tag{9}$$

[2]The information pertaining to U.S. Treasury bills can be sourced from the official website of the United States Department of the Treasury, which is accessible in https://home.treasury.gov/policy-issues/financing-the-government/interest-rate-statistics.

[3]The S&P 500 Index, short for the Standard Poor's 500 Index, is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States.The S&P 500 Index can be utilized as a proxy for the overall performance of market returns.

[4]In this paper, we collect 4-day data before j-th day to compute the expected rate of returns.

	in-stock			cross-stock		
	train	val	test	train	val	test
# of stocks	87	87	87	69	9	9
# of instances	34,730	4,395	4,309	33,354	4,536	5,544
# of tokens	1,482,359	190,716	201,066	1,629,813	113,364	134,004

Table 1: The statistics of in-stock and cross-stock benchmarks

notation	description
X	a sequence of historical prices
x_i	a price on the i-th day
T	a sequence of historical comments
t_i	the comments on the i-th day
Δd	a time window of days
V_j	a price volatility on the j-th day
R_i	a return rate of a stock on the i-th day
$\bar{R}$	a average return rate of a stock
$\bar{R}_m$	a average return rate of stock market
σ_m	a variance of market and stock rate of returns
β_j	a beta coefficient on the j-th day
$\beta_{j-4:j}$	a beta coefficient computed in time interval $[j-4, j]$
D_j	a yield deviation on the j-th day
E_j	a expected rate of return on j-th day
R_f	a risk-free rate of return
M_j	a maximum drawdown on the j-day
r_j	a quantitative investment risk on the j-day
D_{num}	a sequence of numerical data
D_{text}	a sequence of textual data

Table 2: Notations

where $N(x)$ is the min-max normalization function. We categorize investment risk into three levels according to the investment risk quantification.

$$\text{Risk Level} = \begin{cases} \text{low} & 0 \leq r_j < 0.2, \\ \text{medium} & 0.2 \leq r_j < 0.4, \\ \text{high} & 0.4 \leq r_j. \end{cases}$$

The symbol notations used in this paper are summarized in Table 2.

4.3 Dataset Overview

Following the previous works, we take daily historical stock data, including opening prices, closing prices, trading volume, adjusted closing prices, the highest prices, the lowest prices, as well as daily investor comments from Twitter regarding the stocks. We employed the dataset curated by Xu and Cohen (2018) to validate the efficacy of our benchmark model. The dataset encompasses stock data derived from 87 distinct stocks, spanning the temporal interval between 1/1/2014 and 12/31/2015, serving as the basis for analyzing stock investment risk.

risk level	# of instances
low risk	27,428
medium risk	13,637
high risk	2,369

Table 3: The distributions over the risk labels in in-stock and cross-stock benchmarks

We design two benchmarks in-stock and cross-stock. In-stock benchmark aims to evaluate the ability of models which can use the historical information of a stock to predict the future risk of the same stock. The training data spans from 1/1/2014 to 7/8/2015, the validation data spans from 8/8/2015 to 19/10/2015, the test data spans from 20/10/2015 to 31/12/2015. Cross-stock benchmark aims to evaluate the ability of models which can use the historical information of stock to predict the future risk of other stocks. We randomly select 69 stocks as training stock, 9 stocks for val and 9 stock for test, all the data spans from 1/1/2014 to 31/12/2015. Table 1 show the statistics of in-stock and cross-stock benchmarks; and Table 3 shows the distribution of instances across different risk levels.

5 Investment Risk Models

We base the investment risk prediction models on pre-trained language models. As show in figure 2, the framework consists of two components, Numerical Modeling and Textual Modeling.

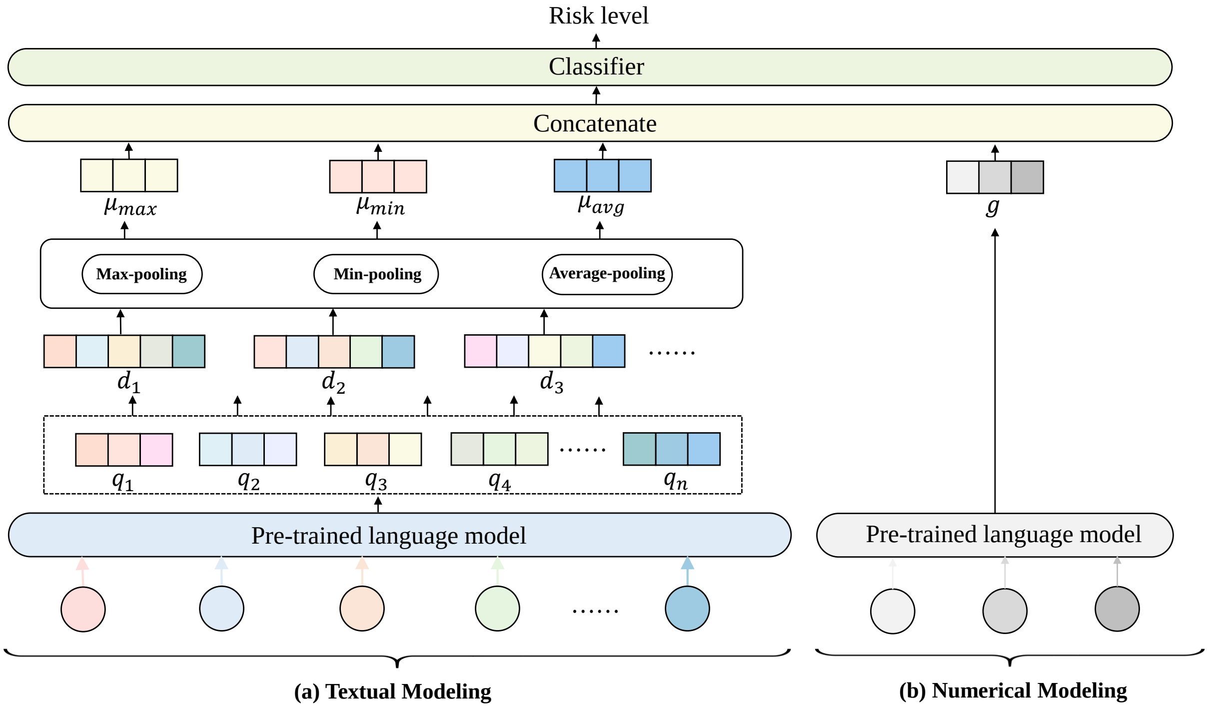


Figure 2: The framework of the the investment risk prediction. (a) Textual modeling obtain the hidden representations of the text features. The maximum pooling, minimum pooling, average pooling is applie and the results are concatenated to obtain the final extracted text feature representation. (b) Numerical modeling obtain the number feature representation g.

5.1 Numerical Modeling

As show in figure 3,we convert numerical data into a textual format with special tokens as identifiers, [OPEN] to denote the opening price, [CLOSE] to indicate the closing price, [HIGH] to indicate the highest price, [LOW] to indicate the lowest price, [ADJ CLOSE] to indicate the adjusted closing price [VOLUME] to indicate the volume. The preprocessed numerical data D_{num} is fed into PLMs to obtain the hidden representations of the corresponding numerical features:

$$g = \text{PLMs}(D_{num}).$$

5.2 Textual Modeling

As show in Figure 3,we select the N latest comments within a time window. Each comment is decorated with timestamp $[\Delta d]$. For example, [0] indicates comments from day d and [-1] indicates comments from day d-1. We input the preprocessed textual data D_{text} into PLMs to obtain the hidden representations of the corresponding textual features. We run the feature extraction as follows:

$$\begin{aligned} q_i &= \text{PLMs}(D_{text}), \\ \mu_{max} &= \text{max-pooling}(q_1, q_2, \ldots, q_n), \\ \mu_{min} &= \text{min-pooling}(q_1, q_2, \ldots, q_n), \\ \mu_{avg} &= \text{average-pooling}(q_1, q_2, \ldots, q_n), \\ C &= [\mu_{max}; \mu_{min}; \mu_{avg}] \end{aligned}$$

where n is the number of comments and q_i denotes hidden representation of i-th comment. We concatenate the outcomes of the maximum, minimum, and average pooling, yielding three distinct feature vectors. The textual representation and numerical representation are concatenated and fed to the risk classifier.

6 Experiments

We conduct experiments on the constructed benchmarks to evaluate the investment risk analysis.

6.1 Experimental Settings

We set a maximum of 30 comments per trading day, with each comment having a maximum length of 128 words, and excess comments and any portion of a comment exceeding the maximum length are

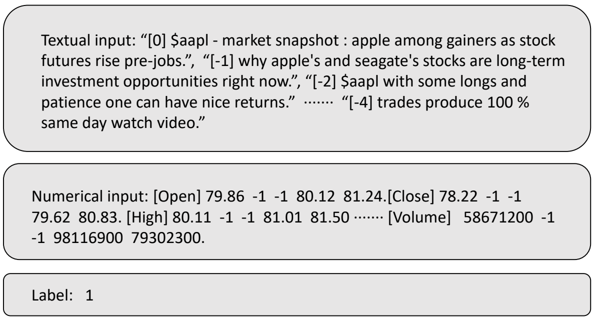


Figure 3: An example of textual inputs and numerical inputs.

truncated. The model are initialized with the pre-trained language model. We use AdamW optimizer (Loshchilov and Hutter, 2017) with a learning rate of 1e-5 in 15 epoches. We employ LoRA (Hu et al., 2021) to fine-tune the models.

6.2 Evaluation metrics

The weighted F1 score is employed as the evaluation metric for our experiments, providing a more nuanced measure of the model's performance across various classes. The weighted F1 score is calculated using the following formula:

$$\mathrm{F1} = \frac{\sum_{i=1}^{n} \alpha_i \cdot \mathrm{F1}_i}{\sum_{i=1}^{n} \alpha_i}$$

where $\mathrm{F1}_i$ is the F1 score for i-th class, α_i is number of the instances in i-th class; and n is the total number of classes.

6.3 Models

We employ different pre-trained language models to validate our benchmark at 2-day and 5-day settings, where 2-day and 5-day means that the input data comes from 2-day and 5-day time window, respectively. Both hidden representation of the numerical features and the textual features are obtained using the same PLM. The PLMs used in the experiment are summarized:

- FinBERT (Araci, 2019) is an enhanced adaptation of the BERT base model, further refined for tasks in financial sentiment analysis.[5]
- RoBERTa (Liu et al., 2019) represents an advanced iteration of the BERT architecture, offering improvements through a more robust training procedure. We employ `RoBERTa-large` in the experiments.[6]
- BERTweet (Nguyen et al., 2020) is an extension of the RoBERTa model, further trained on a corpus derived from Twitter data, improving its efficacy in handling the distinctive linguistic features present in social media text. We employ `BERTweet-large` in the experiments.[7]

model	in-stock 2-day	in-stock 5-day	cross-stock 2-day	cross-stock 5-day
FinBERT	**68.07**	67.63	62.58	63.58
RoBERTa	67.80	69.28	**64.58**	**65.01**
BERTweet	67.82	**69.75**	63.17	64.76

Table 4: Results (F1 %) of the in-stock and cross-stock benchmarks across models.

	in-stock	cross-stock
RoBERTa	**69.28**	**65.10**
w/o textual model	63.80	64.27
w/o numerical model	65.08	56.94

Table 5: Results (F1 %) of RoBERTa without textual model or numerical model.

6.4 Results

In-stock Experiments Table 4 shows the results of the in-stock benchmark and cross-stock benchmark across various models. In the in-stock experiments, FinBERT achieved the highest F1 score of 68.07 within a 2-day time window, while RoBERTa performed better with an F1 score of 69.28 within a 5-day time window. BERTweet achieved the best result in the 5-day settings, benefiting from a larger window that includes more tweet comments suitable for its modeling capabilities.

Cross-stock Experiments In the cross-stock experiments, RoBERTa achieved the highest F1 scores of 64.58 and 65.01 within the 2-day and 5-day time windows, respectively.

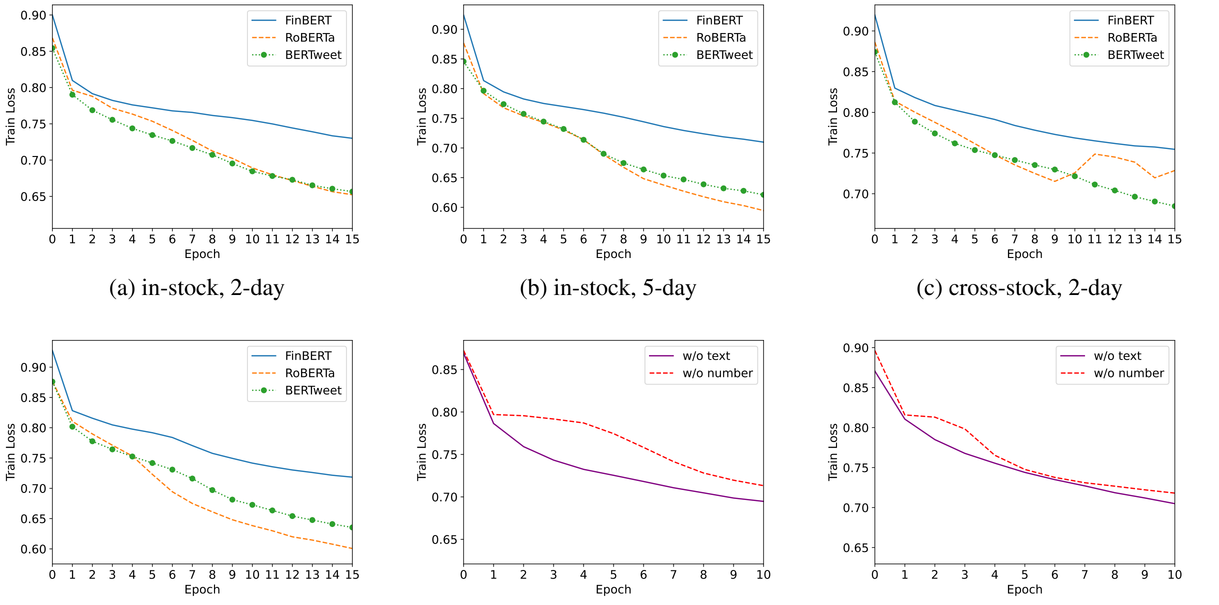
6.5 Analysis

Ablation Study Table 5 shows the results of the ablation experiments with 5-day time window by

[5] https://huggingface.co/ProsusAI/finbert

[6] https://huggingface.co/FacebookAI/roberta-large

[7] https://huggingface.co/vinai/bertweet-large



(a) in-stock, 2-day (b) in-stock, 5-day (c) cross-stock, 2-day

(d) cross-stock, 5-day (e) in-stock ablation (f) cross-stock ablation

Figure 4: The curves of the training loss. (a) is 2-day window training loss in in-stock experiments. (b) is 5-day window training loss in in-stock experiments. (c) is 2-day window training loss in cross-stock experiments. (d) is 5-day window training loss in cross-stock experiments. (e) is the training loss of ablations in RoBERTa in in-stock experiments. (f) is the training loss of ablations in RoBERTa in cross-stock experiments

RoBERTa with only numerical data and only textual data, respectively. The models without textual modeling or numerical modeling have significant decline.

- In in-stock benchmark, textual modeling (-5.48 F1) is more helpful for the investment risk prediction compared to numerical modeling (-4.2 F1). In in-stock benchmark, textual information of a stock in test data is more related to the training data because the historical information of the stocks is in the training data though they are in different periods. This allows the model can capture the link between investment risk and sentiment information more comprehensively. The model exhibits more sensitivity to the embedded sentiment information than the number data, resulting in better predictive performance in investment risk with textual data than the numerical data.

- In cross-stock benchmark, numerical modeling (-8.16 F1) contributes more to the investment risk prediction compared to textual modeling (-0.83 F1). In cross-stock benchmark, the textual information of the stock has relatively weak connections among different stocks. The model is trained to pay more attention to the numerical modeling, resulting that they exhibits more sensitivity with numerical data than the textual data, lead to the better predictive performance in investment risk with numerical data than the textual data.

Training Loss Figure 4 (e) and (f) shows the curves of training losses for RoBERTa ablations. The curves without numerical modeling decline more slowly than the curves without textual modeling, and the without numerical modeling, the training seems to be slightly unstable.

7 Conclusion

We release the novel dataset QuanSIRA for quantitative stock investment risk analysis and propose the in-stock benchmark and the cross-stock benchmark for evaluating and designing investment risk prediction models. Based on the benchmarks, we investigate the application of large language models to tackle stock investment risk analysis. The experimental results show that the quantitative investment risk modeling is challenging, and the performance of the models built on pre-trained language models are marginally acceptable. All the dataset and codes are released to contribute to the stock market analysis in natural language processing area.

Limitations

The quantitative investment risk analysis in this work is only suitable for U.S. stock market, because we use the indexes of U.S. stock. Therefore, the quantification of investment risk can not be applied to other stock markets.

Ethnic Statement

The dataset constructed in this work is baesd on the previous existing dataset. We do not think that this work increases the biases already present in the datasets. Therefore, we do not foresee any ethical issues arising from this work.

References

Ghulam Abbas, Shawkat Hammoudeh, Syed Jawad Hussain Shahzad, Shouyang Wang, and Yunjie Wei. 2019. Return and volatility connectedness between stock markets and macroeconomic factors in the g-7 countries. *Journal of Systems Science and Systems Engineering*, 28:1–36.

Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Zvi Bodie and Alex Kane. 2020. Investments.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

Richard A Brealey, Stewart C Myers, and Franklin Allen. 2014. *Principles of corporate finance*. McGraw-hill.

Jia Chen, Tao Chen, Mengqi Shen, Yunhai Shi, Dongjing Wang, and Xin Zhang. 2022. Gated three-tower transformer for text-driven stock market prediction. *Multimedia Tools and Applications*, 81(21):30093–30119.

Francesco Colasanto, Luca Grilli, Domenico Santoro, and Giovanni Villani. 2022. Albertino for stock price prediction: a gibbs sampling approach. *Information Sciences*, 597:341–357.

Keith Cortis, André Freitas, Tobias Daudert, Manuela Huerlimann, Manel Zarrouk, Siegfried Handschuh, and Brian Davis. 2017. Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. In *Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)*, pages 519–535.

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*.

Xiyong Dong, Youlin Xiong, Siyue Nie, and Seong-Min Yoon. 2023. Can bonds hedge stock market risks? green bonds vs conventional bonds. *Finance Research Letters*, 52:103367.

Yingzhe Dong, Da Yan, Abdullateef Ibrahim Almudaifer, Sibo Yan, Zhe Jiang, and Yang Zhou. 2020. Belt: A pipeline for stock price prediction using news. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 1137–1146. IEEE.

Zihan Dong, Xinyu Fan, and Zhiyuan Peng. 2024. Fnspid: A comprehensive financial news dataset in time series. *arXiv preprint arXiv:2402.06698*.

Frank J Fabozzi and Francesco A Fabozzi. 2021. *Bond markets, analysis, and strategies*. MIT Press.

Eugene F Fama. 1970. Efficient capital markets. *Journal of finance*, 25(2):383–417.

Eugene F Fama and Kenneth R French. 1992. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Saeede Anbaee Farimani, Majid Vafaei Jahan, Amin Milani Fard, and Gholamreza Haffari. 2021. Leveraging latent economic concepts and sentiments in the news for market prediction. In *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–10. IEEE.

Kamaladdin Fataliyev, Aneesh Chivukula, Mukesh Prasad, and Wei Liu. 2021. Stock market analysis with text data: A review. *arXiv preprint arXiv:2106.12985*.

Phil Griffiths. 2016. *Risk-based auditing*. Routledge.

Hemendra Gupta and Rashmi Chaudhary. 2023. An analysis of volatility and risk-adjusted returns of esg indices in developed and emerging economies. *Risks*, 11(10):182.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Esra Ilbahar, Cengiz Kahraman, and Selcuk Cebi. 2022. Risk assessment of renewable energy investments: A modified failure mode and effect analysis based on prospect theory and intuitionistic fuzzy ahp. *Energy*, 239:121907.

Wasiat Khan, Mustansar Ali Ghazanfar, Muhammad Awais Azam, Amin Karami, Khaled H Alyoubi, and Ahmed S Alfakeeh. 2022. Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–24.

Menggang Li, Wenrui Li, Fang Wang, Xiaojun Jia, and Guangwei Rui. 2021. Applying bert to analyze investor sentiment in stock market. *Neural Computing and Applications*, 33:4663–4676.

Tao Li, Feng Ma, Xuehua Zhang, and Yaojie Zhang. 2020. Economic policy uncertainty and the chinese stock market volatility: Novel evidence. *Economic Modelling*, 87:24–33.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Zhuang Liu, Degen Huang, Kaiyu Huang, Zhuang Li, and Jun Zhao. 2021. Finbert: A pre-trained financial language representation model for financial text mining. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, pages 4513–4519.

Alejandro Lopez-Lira. 2023. Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Bernhard Lutz, Nicolas Pröllochs, and Dirk Neumann. 2018. Sentence-level sentiment analysis of financial news using distributed text representations and multi-instance learning. *arXiv preprint arXiv:1901.00400*.

Nagaraj Naik and Biju R Mohan. 2021. Stock price volatility estimation using regime switching technique-empirical study on the indian stock market. *Mathematics*, 9(14):1595.

Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. BERTweet: A pre-trained language model for English Tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 9–14.

Markus Pelger. 2020. Understanding systematic risk: A high-frequency approach. *The Journal of Finance*, 75(4):2179–2220.

Haritha PH and Abdul Rishad. 2020. An empirical examination of investor sentiment and stock market volatility: evidence from india. *Financial Innovation*, 6(1):34.

P Remy and X Ding. 2015. Financial news dataset from bloomberg and reuters.

Raj Shah, Kunal Chawla, Dheeraj Eidnani, Agam Shah, Wendi Du, Sudheer Chava, Natraj Raman, Charese Smiley, Jiaao Chen, and Diyi Yang. 2022. When flue meets flang: Benchmarks and large pretrained language model for financial domain. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2322–2335.

Alan C Shapiro and Paul Hanouna. 2019. *Multinational financial management*. John Wiley & Sons.

William F Sharpe. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442.

Ankur Sinha, Satishwar Kedas, Rishu Kumar, and Pekka Malo. 2022. Sentfin 1.0: Entity-aware sentiment analysis for financial news. *Journal of the Association for Information Science and Technology*, 73(9):1314–1335.

Gaurang Sonkavde, Deepak Sudhakar Dharrao, Anupkumar M Bongale, Sarika T Deokate, Deepak Doreswamy, and Subraya Krishna Bhat. 2023. Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*, 11(3):94.

Priyank Sonkiya, Vikas Bajpai, and Anukriti Bansal. 2021. Stock price prediction using bert and gan. *arXiv preprint arXiv:2107.09055*.

T Swathi, N Kasiviswanath, and A Ananda Rao. 2022. An optimal deep learning-based lstm for stock price prediction using twitter sentiment analysis. *Applied Intelligence*, 52(12):13675–13688.

Jean Tirole. 2010. *The theory of corporate finance*. Princeton university press.

Alessio Venturini. 2022. Climate change, risk factors and stock returns: A review of the literature. *International Review of Financial Analysis*, 79:101934.

Jujie Wang, Jing Liu, and Weiyi Jiang. 2024. An enhanced interval-valued decomposition integration model for stock price prediction based on comprehensive feature extraction and optimized deep learning. *Expert Systems with Applications*, 243:122891.

Neng Wang, Hongyang Yang, and Christina Wang. 2023. Fingpt: Instruction tuning benchmark for open-source large language models in financial datasets. In *NeurIPS 2023 Workshop on Instruction Tuning and Instruction Following*.

Dechun Wen, Tianlong Zhao, Lexin Fang, Caiming Zhang, and Xuemei Li. 2024. Mwdinet: A multilevel wavelet decomposition interaction network for stock price prediction. *Expert Systems with Applications*, 238:122091.

Fenghua Wen, Zhen Liu, Zhifeng Dai, Shaoyi He, and Wenhua Liu. 2022. Multi-scale risk contagion among international oil market, chinese commodity market and chinese stock market: A modwt-vine quantile regression approach. *Energy Economics*, 109:105957.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance.

Qianqian Xie, Weiguang Han, Xiao Zhang, Yanzhao Lai, Min Peng, Alejandro Lopez-Lira, and Jimin Huang. 2023. Pixiu: A large language model, instruction data and evaluation benchmark for finance.

Yumo Xu and Shay B. Cohen. 2018. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1970–1979, Melbourne, Australia. Association for Computational Linguistics.

Yi Yang, Yixuan Tang, and Kar Yan Tam. 2023. Investlm: A large language model for investment using financial domain instruction tuning. *arXiv preprint arXiv:2309.13064*.

Yi Yang, Mark Christopher Siy Uy, and Allen Huang. 2020. Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.

Gafurdjan Zakhidov. 2024. Economic indicators: tools for analyzing market trends and predicting future performance. *International Multidisciplinary Journal of Universal Scientific Prospectives*, 2(3):23–29.

Feng Zhao, Xinning Li, Yating Gao, Ying Li, Zhiquan Feng, and Caiming Zhang. 2022a. Multi-layer features ablation of bert model and its application in stock trend prediction. *Expert Systems with Applications*, 207:117958.

Yu Zhao, Huaming Du, Ying Liu, Shaopeng Wei, Xingyan Chen, Fuzhen Zhuang, Qing Li, and Gang Kou. 2022b. Stock movement prediction based on bi-typed hybrid-relational market knowledge graph via dual attention networks. *IEEE Transactions on Knowledge and Data Engineering*.

Jinan Zou, Qingying Zhao, Yang Jiao, Haiyao Cao, Yanxi Liu, Qingsen Yan, Ehsan Abbasnejad, Lingqiao Liu, and Javen Qinfeng Shi. 2022. Stock market prediction via deep learning techniques: A survey. *arXiv preprint arXiv:2212.12717*.