LLMs for Time Series and Contextual Financial Statement Analysis

Jose Luis Tejada
jtejadac@nd.edu

Kaixiang Zhao
kzhao5@nd.edu

Zixuan Weng
zweng2@nd.edu

1 Introduction

Problem Setting. The field of financial analysis has traditionally relied heavily on quantitative methods, with time series fundamental analysis playing a crucial role in decision-making and research. However, despite the abundance of qualitative data available to financial institutions, there is a significant gap in leveraging qualitative data sources to the same extent of quantitative models. Furthermore, there is a pressing need for models that effectively integrate both quantitative and qualitative data factors to improve predictive accuracy and enhance decision-making.

Motivation. This paper aims to bridge this gap by exploring the potential of large language models (LLMs) for financial statement analysis. By integrating numerical time series analysis with contextual information related to individual company entities and macroeconomic factors, we seek to demonstrate the capabilities of LLMs in performing financial analysis comparable to that of a human financial analyst.

2 Related works

The application of machine learning and artificial intelligence to financial forecasting has seen significant developments in recent years, from traditional statistical models to deep learning architectures and large language models.

Traditional Models. Financial forecasting has its roots in traditional statistical models, with ARIMA Nelson (1998) serving as a foundational approach. Subsequently, with the advent of deep neural networks, Recurrent Neural Networks(RNN), specifically Long Short-Term Memory(LSTM) and Gated Recurrent Units(GRU) tailored for time-series forecasting tasks [Medsker et al. (2001); Shen et al. (2018); Kim & Kang (2019)].

Transformer-based Models. As research progressed, the limitations of RNNs in handling long-range dependencies led to the adoption of Transformer-based architectures. Models like Autoformer Wu et al. (2021), Informer Zhou et al. (2021), and FEDFormer Zhou et al. (2022) have significantly advanced the field by introducing innovative attention mechanisms and feature extraction strategies. PatchTST Nie et al. (2022) refined this approach by applying a patch-based method inspired by computer vision to time series data.

LLMs in Financial Forecasting and Analysis. Large Language Models(LLMs) have recently been applied to financial forecasting and analysis, using their broad knowledge base to interpret time series data. Promptcast Xue & Salim (2023) and LLMTime Gruver et al. (2024) introduced methods to convert numerical data into text, enabling LLMs to perform financial predictions. Google's TEMPO Cao et al. (2023) enhanced forecasting accuracy by incorporating time series decomposition and soft prompts. LSTPrompt Liu et al. (2024a) developed a long short term prompting strategy, further demonstrating LLM's potential as zero-short time series forecasters. Fine-tuning approaches have also shown progress. GPT4TS Zhou et al. (2023) proposed a framework using partially frozen LLMs, achieving competitive performance across various time series analysis tasks. Time-LLM Jin et al. (2023) introduced a method to reprogram time series data, showing strong performance in few-shot and zero-shot settings. In financial statement analysis, Kim & Kang (2019) demonstrated that LLMs can outperform professional analysts in predicting earnings changes without narrative or industry-specific information. Yu et al. (2023) explored LLMs for explainable financial time series forecasting. Surveys by Zhang et al. (2024) and Zhao et al. (2024) have summarized LLM applications in finance, including financial report generation, market trend forecasting, and investor sentiment analysis.

Differences from Other Related Works. We combine time series model with LLM into a unified architecture, using the strengths of both. This integration addresses the limitations of using LLMs or time series models independently. LLMs good at understanding trends but struggle with precise numerical predictions Kim & Kang (2019), while time series models can predict specific values but lack the reasoning capabilities of LLMs. Our model is trained on a custom-built dataset that includes both contextual and numerical data, allowing it to learn the relationships between these data types. Unlike methods that focus solely on financial statements or news, our approach provides multi-data modalities input and simulates real-world conditions by using only data available up to the financial statement release date.

3 DATASETS

To enable our model to effectively incorporate diverse data modalities, we have constructed an extensive dataset that encompasses a wide range of numerical data types, as well as textual datasets. The numerical dataset will include financial statements, ratios calculated from these statements, and macroeconomic metrics. The linguistic dataset will include financial news, as well as investor's analysis.

Financial Statements. The financial dataset was obtained from SEC Edgar financial statement data set, which includes the company balance sheet, income statement and statement of cash flows. The data is provided quarterly since January 2009 to June 2024, which is the most recent dataset as of the writing of this proposal. SEC (January 2009 - June 2024). The SEC provides this data set using eXtensible Business Reporting Language (XBRL) which divides the dataset amongst many disjoint tables SEC (2024). In order to provide the Large Language model with a single set of tables we will use the following helper tool to process the dataset into a single data frame HansjoergW (2024).From this statement we will then use the following formulas to calculate a comprehensive set of financial ratios that will be provided. From this we will be able to create a dataset similar to that used in Kim et al. (2024).

Macroeconomic Data. The OECD Main Economic Indicators database includes a wide range of areas from 1961, for OECD countries and non-member economies. iLibrary (2024) This dataset will provide the necessary macroeconomic contextual information, including items such as quarterly national accounts, business surveys, retail sales, industrial production, construction, consumer prices, total employment, unemployment rates, interest rates, money and domestic finance, foreign finance, foreign trade, and balance of payments. In the case of this research paper, we will only utilize macroeconomic data for the United States.

Financial News. We will be using the FNSPID dataset which is a comprehensive time series financial news dataset. Dong et al. (2024) It is a comprehensive financial dataset designed to enhance stock market predictions by combining quantitative and qualitative data. It contains 29.7 million stock prices and 15.7 million financial news records for 4,775 S&P500 companies from 1999 to 2023. This will provide the model a wide spanning set of market trend, investor and consumer sentiment information that the model will use to contextualize the financial information it is being presented.

4 PROBLEM FORMULATION

The time series forecasting aims to learn a mapping function $\mathcal{F}(\cdot)$ that forecasts the temporal evolution of N variate in the future S time steps based on the observations in the past T time steps and corresponding context information.

$$\mathcal{F}: (\mathbb{R}^{N\times T}, \mathbb{R}^{N\times C}) \to \mathbb{R}^{N\times S},\ (\mathbf{X}, \mathbf{X}_{context}) \mapsto \mathbf{Y} = (\mathbf{x}_{T+1}, \dots, \mathbf{x}_{T+S}),$$

where $\mathbf{x}_t = (c_t^1, \dots, c_t^N) \in \mathbb{R}^N$ denotes the states of N variates at time step t. $\mathbf{X}_{\text{context}}$ is the additional contextual information with dimensions $N \times D$, representing D contextual factors during the prediction period, such as external or macroeconomic factors. Our target variable $\mathbf{Y}$ will consist of the values presented in the financial statement dataset. We do this to avoid focusing the predictive abilities of the model towards a single variable at the expense of others, thus allowing us to showcase the model's holistic understanding.

5 EVALUATION METHODS

Baselines. We compare with the SOTA time series models. Our baselines include **(1): Large Language-Based models:** UniTime (Liu et al. (2024b)), TimeLLM (Jin et al. (2023)), FTP (Zhou et al. (2023)). **(2) Transformer-based models:** iTransformer (Liu et al. (2023)), PatchTST (Nie et al. (2022)), Non-stationary Transformer (Liu et al. (2022b)), Autoformer (Wu et al. (2021)), Informer (Zhou et al. (2021)). **(3) Linear-based models:** DLinear (Zeng et al. (2023)), RLinear (Li et al. (2023)).

Metric details. Regarding metrics, we utilize the mean square error (MSE) and mean absolute error (MAE) for long-term forecasting. In the case of short-term forecasting, we follow the metrics of SCINet (Liu et al., 2022a) on the PeMS datasets, including mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE). As for the M4 datasets, we follow the methodology of N-BEATS (Oreshkin et al., 2019) and implement the symmetric mean absolute percentage error (SMAPE), mean absolute scaled error (MASE), and overall weighted average (OWA) as metrics. It is worth noting that OWA is a specific metric utilized in the M4 competition. The calculations of these metrics are:

$$\text{RMSE} = (\sum_{i=1}^{F}(\mathbf{X}_i - \widehat{\mathbf{X}}_i)^2)^{\frac{1}{2}}, \qquad \text{MAE} = \sum_{i=1}^{F}|\mathbf{X}_i - \widehat{\mathbf{X}}_i|,$$

$$\text{SMAPE} = \frac{200}{F}\sum_{i=1}^{F}\frac{|\mathbf{X}_i - \widehat{\mathbf{X}}_i|}{|\mathbf{X}_i| + |\widehat{\mathbf{X}}_i|}, \qquad \text{MAPE} = \frac{100}{F}\sum_{i=1}^{F}\frac{|\mathbf{X}_i - \widehat{\mathbf{X}}_i|}{|\mathbf{X}_i|},$$

$$\text{MASE} = \frac{1}{F}\sum_{i=1}^{F}\frac{|\mathbf{X}_i - \widehat{\mathbf{X}}_i|}{\frac{1}{F-s}\sum_{j=s+1}^{F}|\mathbf{X}_j - \mathbf{X}_{j-s}|}, \qquad \text{OWA} = \frac{1}{2}\left[\frac{\text{SMAPE}}{\text{SMAPE}_{\text{Naïve2}}} + \frac{\text{MASE}}{\text{MASE}_{\text{Naïve2}}}\right],$$

where s is the periodicity of the data. $\mathbf{X}, \widehat{\mathbf{X}} \in \mathbb{R}^{F\times C}$ are the ground truth and prediction results of the future with F time pints and C dimensions. $\mathbf{X}_i$ means the i-th future time point. We will evaluate the performance of the model by comparing its predicted financial statement values to the actual financial statement values being reported for the unobserved time period(s. We will evaluate the performance of the model for both short and long-term predictive abilities using the respective metrics above.

6 MILESTONES

In order to ensure the proper progress of the project, we have created the following milestones focused on completing simple measurable steps towards the final goal of creating a LLM model capable of performing company analysis like a human. For each of these milestones we will schedule a discussion with the professor for the following Monday.

Table 1: Milestones Table

Date	Main Task and Description
October 12	**Preliminary Analysis of Companies** Perform simple analysis of 5 companies with a dataset of 3 years of data, predicting values for the next quarter. Perform time series LLM reprogramming. Compare the predictive performance of the model against using the values of the previous time period and the average value over selected time periods.
October 28	**Deeper Company Analysis** Continue analyzing 5 companies, increase dataset size. **If LLM works better than baseline:** Expand the data size to include 15 years worth of data. Compare the results with those achieved by other works. **If LLM does not work better than baseline:** Use traditional time series model to provide numerical predictions and then task the LLM with modifying those predictions to reduce the overall error against the actual predicted values.
November 21	**Full Dataset and Compare Against Traditional Time Series Models** Use the full dataset from 2009 Q1 to 2024 Q3, analyze a total of the 500 largest companies in the US (S&P 500). Compare the performance of our model against established time series models listed in related works.
December 6	**Complete Project and Paper** Write entire paper and report results. Send full paper to Professor Jiang to obtain feedback and revise as necessary.

7 CONCLUSION

This proposal presents a novel approach to financial forecasting by combining time series models with large language models. The suggested method aims to integrate financial statement and contextual data in both numerical and textual forms. We anticipate that this approach could improve the capture of complex financial relationships and generate accurate predictions.

REFERENCES

Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint arXiv:2310.04948*, 2023.

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

Nate Gruver, Marc Finzi, Shikai Qiu, and Andrew G Wilson. Large language models are zero-shot time series forecasters. *Advances in Neural Information Processing Systems*, 36, 2024.

HansjoergW. sec-fincancial-statement-data-set, 2024.

OECD iLibrary. Main economic indicators, 2024.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*, 2023.

Alex Kim, Maximilian Muhn, and Valeri Nikolaev. Financial statement analysis with large language models. *arXiv preprint arXiv:2407.17866*, 2024.

Sangyeon Kim and Myungjoo Kang. Financial series prediction using attention lstm. *arXiv preprint arXiv:1902.10877*, 2019.

Zhe Li, Shiyi Qi, Yiduo Li, and Zenglin Xu. Revisiting long-term time series forecasting: An investigation on linear mapping. *arXiv preprint arXiv:2305.10721*, 2023.

Haoxin Liu, Zhiyuan Zhao, Jindong Wang, Harshavardhan Kamarthi, and B Aditya Prakash. Lstprompt: Large language models as zero-shot time series forecasters by long-short-term prompting. *arXiv preprint arXiv:2402.16132*, 2024a.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. SCINet: time series modeling and forecasting with sample convolution and interaction. *NeurIPS*, 2022a.

Xu Liu, Junfeng Hu, Yuan Li, Shizhe Diao, Yuxuan Liang, Bryan Hooi, and Roger Zimmermann. Unitime: A language-empowered unified model for cross-domain time series forecasting. In *Proceedings of the ACM on Web Conference 2024*, pp. 4095–4106, 2024b.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in Neural Information Processing Systems*, 35:9881–9893, 2022b.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023.

Larry R Medsker, Lakhmi Jain, et al. Recurrent neural networks. *Design and Applications*, 5(64-67): 2, 2001.

Brian K Nelson. Time series analysis using autoregressive integrated moving average (arima) models. *Academic emergency medicine*, 5(7):739–744, 1998.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *ICLR*, 2019.

SEC. Financial statement data set documentation, 2024.

SEC. Financial statement data set, January 2009 - June 2024.

Guizhu Shen, Qingping Tan, Haoyu Zhang, Ping Zeng, and Jianjun Xu. Deep learning with gated recurrent unit networks for financial sequence predictions. *Procedia computer science*, 131:895–903, 2018.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

Hao Xue and Flora D Salim. Promptcast: A new prompt-based learning paradigm for time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

Xinli Yu, Zheng Chen, Yuan Ling, Shujing Dong, Zongyi Liu, and Yanbin Lu. Temporal data meets llm–explainable financial time series forecasting. *arXiv preprint arXiv:2306.11025*, 2023.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11121–11128, 2023.

Xiyuan Zhang, Ranak Roy Chowdhury, Rajesh K Gupta, and Jingbo Shang. Large language models for time series: A survey. *arXiv preprint arXiv:2402.01801*, 2024.

Huaqin Zhao, Zhengliang Liu, Zihao Wu, Yiwei Li, Tianze Yang, Peng Shu, Shaochen Xu, Haixing Dai, Lin Zhao, Gengchen Mai, et al. Revolutionizing finance with llms: An overview of applications and insights. *arXiv preprint arXiv:2401.11641*, 2024.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, pp. 27268–27286. PMLR, 2022.

Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36:43322–43355, 2023.