

Deep learning model-driven financial risk prediction and analysis

Tianyi Yang^{1a,*}, Ang Li^{1b}, Jiahao Xu², Guangze Su³, Jiufan Wang⁴

^{1a}Financial Risk Management, University of Connecticut, Stamford CT, USA

^{1b}Business Analytics, University College Dublin, Dublin, Ireland

²Master of Science in Financial Engineering, University of Southern California, CA, USA

³Information Studies, Trine University, Phoenix, USA

⁴Independent Researcher, William & Mary, Williamsburg, VA, USA

*Corresponding author E-mail:donaldyan3@gmail.com

Abstract. The integration of deep learning models into financial risk prediction and analysis has significantly transformed traditional approaches. While conventional quantitative methods often rely on simplistic metrics like maximum drawdown, the advent of deep learning necessitates a more nuanced evaluation, emphasizing the model's generalization ability, especially during market crises such as stock market crashes. This paper explores the critical aspects of evaluating deep learning models' risk control capabilities in finance, underscoring the importance of understanding both statistical metrics and generalization abilities, particularly in adverse market conditions. The experimentation reveals that different deep generative models excel in various aspects of financial time series analysis, with generative adversarial networks (GANs) demonstrating superior performance in predicting Value at Risk (VaR) and variational autoencoders (VAEs) excelling in return rate prediction. Moreover, integrating multiple models further enhances predictive performance, leveraging the strengths of each model to compensate for individual weaknesses. Overall, this paper underscores the potential and significance of deep generative models in financial time series analysis, offering a roadmap for improved risk management and decision-making in financial markets.

Keywords: Deep learning models, Financial risk prediction, Generative adversarial networks (GANs), Value at Risk (VaR), Model integration.

1. Introduction

In today's dynamic financial landscape, the integration of deep learning models has revolutionized the approach to risk prediction and analysis. Traditional quantitative methods often rely on metrics like maximum drawdown to evaluate risk control capabilities. However, with the advent of deep learning, there's a growing recognition that risk assessment requires a more nuanced understanding, extending beyond simplistic measures. Deep learning models offer unparalleled sophistication in capturing complex patterns and dynamics within financial data. [1]A model may perform exceptionally well in backtesting or even during a limited live trading period. Still, its true test lies in its response to unprecedented market turmoil.

This article delves into the critical considerations for evaluating the risk control capabilities of deep learning models in finance. It underscores the importance of not only statistical metrics but also a comprehensive understanding of a model's generalization ability, particularly in the face of adverse market conditions. By exploring how deep learning models fare in scenarios such as stock market crashes and emphasizing the significance of cross-validation techniques, this article aims to provide practitioners with insights into building robust risk management systems. Ultimately, it advocates for a holistic approach that combines quantitative analysis with an understanding of macroeconomic factors to enhance financial risk prediction and analysis in today's volatile markets.

2. Related work

2.1. Deep learning

The origins of deep learning can be traced back to the 1950s, and Bloom's classification of cognitive dimensions in the Classification of Educational Goals reflects the idea that there are deep and shallow learning levels. In "Efficient deep learning: A survey on making deep learning models smaller, faster, and better," the author explores strategies and techniques aimed at optimizing deep learning models for improved efficiency[2]. The survey provides insights into methods for reducing model size, enhancing speed, and boosting overall performance, catering to the growing demand for more efficient deep learning algorithms. "Machine learning and deep learning," published in Electronic Markets, delves into the evolving landscape of machine learning and deep learning methodologies. [3] The article discusses advancements, challenges, and applications in these fields, highlighting their significance across various domains such as business and technology. Presented at the International Conference on Global Research and Education, "List of deep learning models" compiles a comprehensive overview of various deep learning architectures. [4] The article serves as a valuable resource for researchers and practitioners in the field, offering insights into the diverse range of models and their contributions to the advancement of artificial intelligence.

It is concluded that deep learning has five characteristics: focusing on the cultivation of critical thinking ability, emphasizing the correlation and integration of information, promoting the construction and reflection of knowledge, the transfer and application of conscious knowledge and ability, and the cultivation of problem-solving[5]. It is also pointed out that these five characteristics of deep learning are not isolated, but interrelated as a whole, which jointly promote the realization of deep learning.

2.2. Deep learning model

Deep learning is an important branch of the field of artificial intelligence, which has made remarkable progress in recent years. Among them, RNN, CNN[6], Transformer, BERT and GPT are five commonly used deep learning models, which have made important breakthroughs in computer vision, natural language processing and other fields. RNN is a kind of neural network model, its basic structure is a cyclic body, which can process sequence data. RNNs are characterized by their ability to process the current input while remembering previous information. This structure makes RNNs well suited for tasks such as natural language processing and speech recognition, which require processing data with temporal relationships.

CNN is a kind of neural network model, its basic structure is composed of multiple convolutional layers and pooling layers. This structure of CNN makes it well suited for computer vision tasks such as image classification, object detection, and so on. Compared with RNN, CNN is better at processing image data because it can automatically learn local features in the image without the need to manually design a feature extractor.

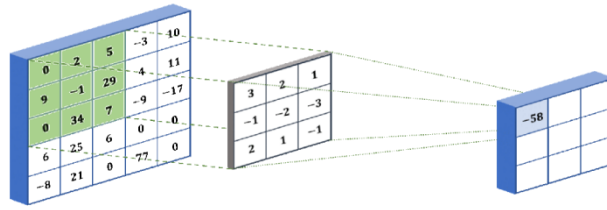


Figure 1. Convolutional network model frame diagram

How convolutional neural networks work:

- CNN applies filters (small rectangles) to the input image to detect features such as edges or shapes. The filter slides over the width and height of the input image and calculates the dot product between the filter and the input to generate the activation graph.
- The activation graph is fed into the pooling layer, and the graph is downsampled to reduce the dimension. This makes the model more efficient and robust. The final layer is the fully connected layer, which classifies the input images into categories such as "dog" or "cat."
- Some popular CNN architectures include AlexNet, VGGNet, ResNet, and Inception. These have been used to solve complex problems, such as identifying thousands of objects or detecting disease through medical scans.

Transformer is a neural network model based on self-attention mechanism, and its basic structure is composed of multiple encoders and decoders. The encoder can convert an input sequence into a vector representation, while the decoder can convert that vector representation back into an output sequence. Transformer's biggest innovation is the introduction of a self-attention mechanism, which allows the model to better capture long-distance dependencies in the sequence. Transformer has seen great success in natural language processing, such as machine translation, text generation and other tasks.

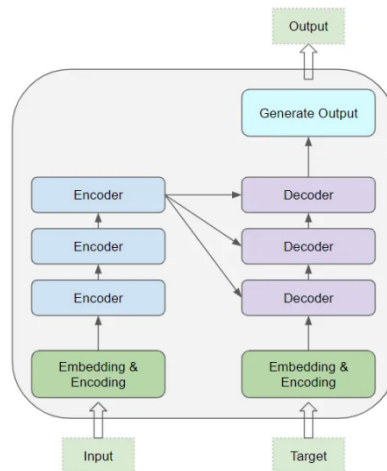


Figure 2. Transformer[7] Multiple encoder frameworks

As can be seen from Figure 2, the core of Transformer is composed of multiple Encoder and Decoder layers. Here, we refer to a single layer as an encoder or decoder, and a group of such layers as an encoder or decoder group. The encoder group and the decoder group each have their respective Embedding layers to process their own input data. Finally, the final result is generated through an output layer.

BERT is a pre-trained language model based on the Transformer architecture, with its key innovation being a bi-directional Transformer encoder that considers contextual information from both directions in the input sequence. BERT is pre-trained on large text datasets to learn extensive language knowledge and is then fine-tuned for specific tasks like text classification and sentiment analysis. It has achieved

significant success in natural language processing (NLP) and is widely used for various NLP tasks. The Transformer architecture enables BERT[8] to process input sequences by simultaneously considering contextual information from all positions.

The advantages of BERT model in forecasting are mainly reflected in the following aspects:

- **Context understanding:** BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model whose biggest feature is the introduction of bidirectional context understanding. BERT is able to consider both the left and right side of the context, better capturing meaning and relationships within the context.
- **Richness of semantic representation:** BERT models are pre-trained on large-scale corpora and learn rich semantic representations.
- **Fine-tuning mechanism:** BERT models can translate general language understanding into task-specific performance improvements through fine-tuning on specific tasks. By fine-tuning domain-specific datasets, BERT can be personalized optimized for different tasks, improving prediction accuracy.

2.3. Risk prediction model

Financial risk prediction is crucial in the financial industry, and machine learning has been widely used to identify potential risks automatically, reducing the need for manual labor. However, recent developments in this field have not kept pace with two key issues:

1. **Outdated Algorithms:** Many current algorithms are outdated, especially given the rapid advancements in generative AI and large language models (LLMs)[9].
2. **Lack of Unified Benchmark:** The absence of a unified and open-source financial benchmark has hindered research progress for years.

To address these issues, we propose two solutions:

FinPT (Financial Profile Tuning): This is a novel approach to financial risk prediction that utilizes the latest advancements in large pre-trained base models, enhancing their effectiveness through profile tuning;
FinBench: This is a high-quality dataset specifically designed for financial risk prediction, including data on default, fraud, and customer churn; Together, FinPT and FinBench aim to modernize financial risk prediction and provide a robust, standardized dataset to facilitate ongoing research and development.

This paper proposes a deep generative modeling method for financial time series to enhance VaR (Value at Risk) estimation. The method uses deep learning techniques to simulate the probability distribution of financial time series, enabling better future risk predictions. It highlights the unique challenges of financial time series, such as nonlinearity, non-normal distribution, and volatility clustering. The paper then explains deep generative modeling, which involves using deep learning to create samples that mimic financial time series characteristics. Finally, it discusses VaR's definition and significance, noting the limitations of traditional VaR[10-11] estimation methods.

3. Methodology

3.1. Data preprocessing

When pre-processing financial time series data, it is usually necessary to perform operations such as cleaning, smoothing and standardization to improve the stability and accuracy of the model. These time Windows are stitched together into a three-dimensional array for training and fitting the model. The first 10 days of each window are used as conditions and the last 10 days are used as targets. Thus, each two-dimensional array can be called a data frame, window, slice, or sequence. This data processing approach not only helps to improve the performance of the model, but also enables the model to better understand the correlations and trends in the time series data. By converting time series data into two-dimensional arrays and leveraging the powerful properties of deep learning models, we can more accurately predict changes in financial markets, thereby better managing risk and making investment decisions.

3.2. Build Deep Learning Model

In this experiment, deep learning models (such as generative adversarial network, variational autoencoder, diffusion model, etc.) are used to build a generative model of financial time series, and new samples are generated by learning the distribution characteristics of data.

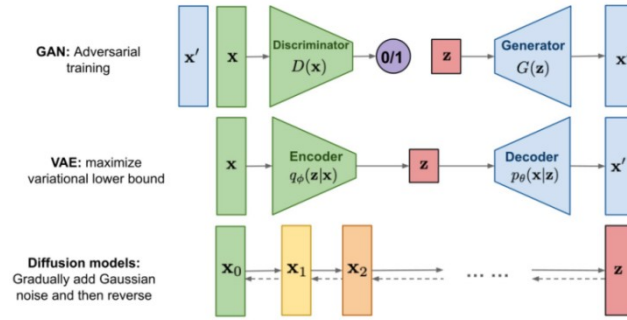


Figure 3. Architecture comparison among three group of generative models

In this experiment, each model has its own unique role and advantages (Figure 3). Generative adversarial networks (GANs) learn the distribution features of data by competitively training generators and discriminators.

Variational autoencoders (VAE) are probabilistic generation models that generate new samples by learning the potential variable space of the data. VAE maps the input data into the potential space via an encoder and the potential variables back into the original data space via a decoder. By modeling the data in the underlying space, VAE is able to generate financial time series data with diversity and continuity to better capture the complex features of the market.

3.3. Model training and optimization

Training the generated model with training data and optimizing the model parameters through optimization algorithms (such as stochastic gradient descent) to improve the fitting ability and generalization ability of the model are key steps. Through continuous iteration of the training data, the generation model can gradually learn the distribution characteristics of the data, so as to generate more realistic samples. The selection of optimization algorithm and parameter adjustment are very important to improve the performance of the model.

In this paper, the empirical research and comparison with traditional methods show that the deep generation modeling method has significant effectiveness and advantages in financial time series analysis. Compared with traditional methods, the deep generation modeling method can capture the nonlinear characteristics and tail risks of financial time series more accurately, and improve the accuracy and stability of risk measurement. The advantages of this approach lie in its better ability to model the distribution of data, as well as the fidelity and diversity of the generated samples, thus providing a more reliable basis for risk management and decision-making in financial markets.

3.4. Experimental design

1. Data sets: This paper uses three financial time series data sets, namely S&P 500, Nikkei 225 and Euro Stoxx 50.

2. Concept and definition: This paper uses deep generation model to model financial time series, including autoencoder, variational autoencoder, generative adversarial network and other models. This paper also introduces the concept and definition of value risk (VaR), which is an important index to measure financial risk.

3. Experimental indicators: This paper introduces a method for comparing the performance of different generation models, and measures the similarity between the generated synthetic data and the real data through quantitative and qualitative indicators.

Qualitative indicators include visual comparison of empirical distribution, t-SNE, PCA and UMAP. The quantitative indexes include distribution distance, ACF and backtest. The authors use short - and long-path composite samples to evaluate model performance and provide a comprehensive score to compare the performance of different models. The model ranking depends on the KPIs used and the way the scores are combined.

3.5. *Experimental result*

In this paper, we compare the performance of different deep generation models on three different financial datasets. We find that generative adversarial networks (Gans) perform best in predicting VaR (Value at Risk). By competitively training generators and discriminators, Gans can effectively capture the features of data distribution and generate high-quality samples. When it comes to predicting returns, we find that variational autoencoders (VAE) perform best. VAE is a probabilistic generation model that generates new samples by learning the potential variable space of the data. Its ability to generate samples with diversity and continuity gives VAE an advantage in predicting the return rate of financial assets. VAE can improve the forecasting accuracy of financial time series data by effectively modeling the rate of return.

In summary, the results of this paper show that different deep generation models have their own advantages and applicable scenarios in financial time series data prediction. By reasonably selecting and integrating multiple models, the forecasting performance can be further improved to provide more reliable support for risk management and decision making in financial markets.

4. Conclusion

The application of deep generation model in the financial field has extensive potential and important significance. With the continuous development and maturity of deep learning technology, deep generation models can better capture complex patterns and dynamic characteristics in financial time series data, and provide new ideas and tools for financial market prediction, risk management and decision-making. Deep generation models can be applied to risk measurement and management in financial markets. In addition to VaR estimates, deep generation models can also be used to predict other important Risk indicators, such as Conditional Value at Risk, Expected Shortfall, etc. Through comparative and empirical studies of different deep-generating models, their performance and applicability to different risk measures can be evaluated more comprehensively.

Deep generation models can also be used to enhance and extend financial time series data. Traditional time series data are often affected by noise and missing values, which limits the performance and predictive power of the model. The simulation generates a large amount of financial time series data, which can evaluate the performance of different strategies and trading rules in different market environments, and provide references for portfolio optimization and risk management. In the future, the deep generation model can also be combined with other deep learning technologies, such as reinforcement learning, transfer learning, etc., to further improve the performance and application range of the model. Through continuous innovation and exploration, the deep generation model will play an increasingly important role in the financial field and provide more powerful support for the stable and healthy development of the financial market.

References

- [1] Menghani, G. (2023). Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Computing Surveys*, 55(12), 1-37.
- [2] Menghani, Gaurav. "Efficient deep learning: A survey on making deep learning models smaller, faster, and better." *ACM Computing Surveys* 55.12 (2023): 1-37.
- [3] Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685-695.

- [4] Mosavi, A., Ardabili, S., & Varkonyi-Koczy, A. R. (2019, September). List of deep learning models. In International conference on global research and education (pp. 202-214). Cham: Springer International Publishing.
- [5] Peng, Kuashuai, and Guofeng Yan. "A survey on deep learning for financial risk prediction." *Quantitative Finance and Economics* 5.4 (2021): 716-737.
- [6] Yu, Jiyeon, Angelica de Antonio, and Elena Villalba-Mora. "Deep learning (CNN, RNN) applications for smart homes: a systematic review." *Computers* 11.2 (2022): 26.
- [7] Hanslo, Ridewaan. "Deep learning transformer architecture for named-entity recognition on low-resourced languages: State of the art results." 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS). IEEE, 2022.
- [8] Justus, D., Brennan, J., Bonner, S., & McGough, A. S. (2018, December). Predicting the computational cost of deep learning models. In 2018 IEEE international conference on big data (Big Data) (pp. 3873-3882). IEEE.
- [9] Koumakis, L. (2020). Deep learning models in genomics; are we there yet?. *Computational and Structural Biotechnology Journal*, 18, 1466-1473.
- [10] Qin, Duo. "Rise of VAR modelling approach." *Journal of Economic Surveys* 25.1 (2011): 156-174.
- [11] Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410).