THE USA JOURNALS
THE AMERICAN JOURNAL OF MANAGEMENT AND ECONOMICS INNOVATIONS (ISSN- 2693-0811)
VOLUME 06 ISSUE12

PUBLISHED DATE: - 18-12-2024
DOI: - https://doi.org/10.37547/tajmei/Volume06Issue12-06 PAGE NO.: - 55-62

RESEARCH ARTICLE Open Access

USING SYNTHETIC DATA TO MODEL A PORTFOLIO IN CONDITIONS OF HIGH VOLATILITY. HOW SYNTHETIC DATA ALLOWS YOU TO TEST STRATEGIES FOR RARE MARKET EVENTS. EXAMPLES OF GENERATIVE MODELS APPLICATION

Zharmagambetov Yernar
Investment portfolio manager, MBA, Almaty, Kazakhstan

Abstract
The article considers the possibility of using synthetic data to model a portfolio in conditions of instability in financial markets. Methods based on the analysis of historical data need to be revised to account for rare events that are missing from historical records. The use of generative models makes it possible to generate synthetic data, creating conditions for modeling situations that go beyond the known scenarios. The purpose of the article is to consider the impact of synthetic data on asset management methods in conditions of high market volatility. The methods of creating synthetic data using generative adversarial networks and their role in modeling situations with limited access to necessary information are described. The use of synthetic data in scientific papers confirms their effectiveness in adapting asset management strategies, which contributes to improving results. This underscores the need for their application and guarantees the stability of financial systems in the external conditions that determine the present. In conclusion, it is noted that generative models that create synthetic data increase the accuracy and flexibility of financial portfolio management strategies. The considered approaches open up opportunities for forecasting and decision-making. Due to this, the information contained in the work will be useful to investors and bank employees.

Keywords Synthetic data, generative models, volatility, rare events, investment strategies, machine learning.

INTRODUCTION

Modern financial markets are characterized by high volatility, limiting the applicability of traditional methods for forecasting and testing investment strategies under extreme fluctuations. In such circumstances, it is essential to develop methods that model asset behavior in rare and unconventional market situations. These methods facilitate accurate assessments of portfolio stability and profitability. One such approach involves the use of synthetic data generated to analyze various market scenarios, including financial crises, sharp asset price fluctuations, and changes in exchange rates.

According to estimates by the McKinsey Global Institute, generative AI could deliver between $2.6 trillion and $4.4 trillion annually across various industries worldwide in the 63 scenarios analyzed. Among industries, the banking sector is expected to realize some of the greatest benefits, with an annual potential ranging from $200 billion to $340 billion (equivalent to 9–15 percent of operating profit), primarily driven by productivity gains. The economic impact is likely to be positive across all banking segments and functions, with the highest absolute benefits projected in the corporate and retail sectors ($56 billion and $54 billion, respectively) [11].

These models replicate dependencies characteristic of real markets, including extraordinary situations. The application of such methods enables the testing of strategies under conditions that are rarely represented in historical data, improving risk forecasting and enhancing portfolio resilience.

The use of synthetic data requires the development of new methods for risk assessment and portfolio optimization under conditions of economic instability. Amid environmental changes and market uncertainty, employing synthetic data for testing and adapting strategies represents a critical task, as it expands modeling capabilities and improves the quality of investment decision-making.

The purpose of this article is to explore the potential of using synthetic data for portfolio modeling under conditions of high volatility, as well as to examine practical examples.

METHODS

Studies dedicated to forecasting rare events highlight the necessity of employing generative adversarial networks (GANs) to create synthetic data. Articles by Labiad B., Berrado A., Benabbou L. [1], and Dannels, S. [4] demonstrate the application of these models for generating market data under conditions of increased volatility. This approach enables the modeling of phenomena that are challenging to capture using traditional methods. Issues related to the scarcity of data on market surges are addressed through conditional generative adversarial networks, as described in the work of Chen Y. et al. [9].

Research by Tepelyan R., Gopal A. [5], Gibson L., Hoerger M., and Kroese D. [6] confirms the effectiveness of these networks in forecasting returns. These methods aid in the development of strategies that account for changing market conditions.

The article by Naritomi Y. and Adachi T. [2] discusses the use of generative networks to improve trading forecasts, while Dogariu M. et al. [3, 10] explore the potential of generative networks for testing strategies.

The method proposed by Li J. et al. [7] focuses on the application of Wasserstein in generating synthetic flows of market orders, facilitating the modeling of market liquidity. Coletta A. et al. [8] use conditional generative adversarial networks to simulate market scenarios, expanding the possibilities for modeling various strategies.

The practical examples described in this study draw on sources [13-14], available on the websites *masterofcode.com* and *www.orientsoftware.com*, which detail the experiences of companies using generative networks for portfolio modeling. Source [11], located on the website *www.mckinsey.com*, provides statistical data related to the use of generative intelligence.

The article employs a method of analyzing existing scientific literature and reviewing real-world examples in the field.

RESULTS AND DISCUSSIONS

Financial markets in recent decades have been characterized by high volatility, complicating asset management. In such conditions, methods that adequately reflect ongoing changes are essential. Forecasts based on historical data analysis lose their effectiveness in predicting future developments. Synthetic data generated using mathematical models offer opportunities for developing risk management techniques.

Synthetic data are artificially created datasets with statistical properties similar to real indicators. They are generated using generative adversarial networks (GANs) and machine learning methods. These technologies enable the modeling of rare market situations that are infrequent in historical data but significantly impact investment decision-making [1, 4, 9].

The use of synthetic data facilitates the study of complex market situations. Traditional methods do not capture the full spectrum of risks, necessitating alternative approaches. Artificial data helps mitigate information deficits by modeling events not accounted for in historical analysis.

Various algorithmic approaches are used to create synthetic data. Generative adversarial networks consist of two neural models: one generates data, while the other evaluates their conformity to predefined parameters. GANs serve as a tool for generating data, including time series of market activity. The model architecture includes a generator that forms information based on random input signals and a discriminator that assesses the degree of similarity [6]. The objective is to achieve a state where the differences between artificially generated and real data become imperceptible.

Below is the formula used for modeling data under conditions of high volatility in financial markets. In the context of portfolio modeling during high volatility, GANs are employed to generate realistic time series that simulate market data, such as stock prices or indices subject to significant fluctuations. The generator attempts to create time series resembling real data, while the discriminator evaluates their similarity.

$$min_G max_D V(D,G) = E_{x\sim Pdata(x)}[log\ log\ D(x)] + E_{z\sim Pz\ (z)}[log\ log\ (1 - D(G(z)))] \quad (1)$$

Where:

- G is the generator, which transforms a random vector z from the distribution $P_z(z)$ into data (e.g., an image).

- D is the discriminator, which evaluates whether an input example is real (from the data distribution $P_{data}(x)$) or generated (from the generator $G(z)$).

- x represents data drawn from the distribution $P_{data}(x)$.

- z represents random data (typically latent variables or noise) fed into the generator.

- $G(z)$ represents the generated data output by the generator when z is provided as input.

The ideal case (equilibrium) is achieved when the generator produces data that the discriminator cannot distinguish from real data, meaning D(x) = 0.5 for all x, and the discriminator cannot differentiate between real and generated examples. This process models a zero-sum game between two agents (the generator and the discriminator), forming the basis for training.

These models are used to create surfaces that reproduce historical data while incorporating elements of market behavior. This approach facilitates stress testing by simulating various crisis scenarios [3,10]. Table 1 below outlines the advantages and disadvantages of using generative networks.

Table 1. Advantages and Disadvantages of Using Generative Networks (compiled by the author)

Advantage	Disadvantage
Generative networks effectively model complex nonlinear relationships between assets, useful for identifying hidden risk factors and market opportunities.	Forecasts can complicate integration and justification, posing challenges for asset managers and investors.
Generative models create synthetic data, including rare or extreme events (e.g., crises), critical for risk assessment in highly volatile conditions.	Reliable training of generative models requires large amounts of high-quality data, which may not be available in volatile conditions.
Generative networks can help simulate portfolios resilient to unexpected market fluctuations by generating numerous alternative scenarios.	Generative models demand significant computational power, increasing the cost of their use.
Generative networks enable the development of new asset allocation strategies that adapt to changing market conditions and high volatility.	Improper tuning or insufficient data can lead to overfitting, resulting in poor generalization and unstable outcomes.
Generative networks are used to predict future market states based on trends and historical data, crucial during periods of instability.	Results are highly dependent on model hyperparameter selection (e.g., network architecture, learning rate), requiring careful tuning and optimization.
Generative models reduce the risk of asset overvaluation or forecasting errors common in traditional methods through more complex and flexible data structures.	Generative models can be sensitive to noise and unreliable data, potentially leading to incorrect conclusions and decisions.

Synthetic data expands the range of analyzed scenarios, including rare market phenomena. This enhances risk assessment accuracy and identifies vulnerabilities in management practices.

Generative models create scenarios that allow testing the resilience of investment strategies in unstable market environments.

Mastercard actively employs algorithms to combat fraud and ensure the protection of customer data. As cyber threats continue to grow, detecting fraudulent activities becomes increasingly challenging. Criminals use methods such as spyware, malware, and skimming to gain access to card data. Many limit their activities to selling fragments of information on dark markets, making detection more difficult.

Algorithms enable financial organizations to efficiently process real-time data, identifying patterns indicative of fraud. Predictive models forecast potential criminal schemes, allowing timely responses to threats. Mastercard has reported [12] that these technologies have doubled the detection rate of compromised cards while reducing false positives in the analysis of suspicious transactions. Additionally, the verification of high-risk merchants has been significantly accelerated.

JPMorgan Chase developed the LLM Suite AI assistant system to support its 60,000 employees [12]. This tool automates routine tasks, enabling employees to focus on more critical aspects of their work. The system's functionality includes protecting data from external threats and training staff in effective technology utilization methods.

However, the bank acknowledges that implementing such technologies will result in changes to workforce structures. Automating certain processes may reduce the number of jobs in specific areas. It is crucial to emphasize that the use of AI must be carefully managed, particularly when handling sensitive data, to prevent the dissemination of incorrect conclusions or misinformation [12].

Forrester reports [13] that nearly 70% of decision-makers in the banking sector believe personalization is essential for effective customer service. These efforts are also supported by company executives, who recognize that a personalized approach is critical for business success. However, only 14% of surveyed consumers currently rate banks as providing excellent personalized service (Figure 1).

Figure 1. Forrester Survey on Generative System Applications in Banking [13]

Generative AI systems help address issues in the banking sector by tailoring services to individual customer characteristics. They analyze goals, preferences, and risk levels, providing recommendations that align with unique conditions. This approach enables the creation of precise offerings, improving service quality and customer satisfaction. The technology responds to current market trends and considers economic forecasts, offering critical data for decision-making.

In investment banking, generative models significantly enhance the processing of large data volumes. They uncover hidden patterns that might otherwise go unnoticed, improving the accuracy of risk assessments and simplifying the decision-making process. Models based on trend information generate forecasts that give analysts a comprehensive understanding of the potential outcomes of various actions. Combining advanced analytics with powerful algorithms creates opportunities to enhance results for all stakeholders, from investors to financial institutions [13].

Table 2. Advantages and Disadvantages of Using Synthetic Data for Portfolio Modeling (compiled by the author)

Advantages	Disadvantages
Synthetic data enable the creation of rare market events, such as crises or significant corrections, which are difficult to observe in real data but are essential for portfolio stress testing.	Generating synthetic data that accurately reflects real market conditions is challenging, particularly when attempting to replicate complex economic and political factors.
The use of synthetic data allows testing strategies under rare but possible extreme market events, such as liquidity crashes or sharp volatility shifts.	Synthetic data do not always fully model all risks associated with real markets, potentially leading to an underestimation of risks during instability.
Generative models facilitate varying market conditions, creating both highly volatile and stable market scenarios necessary for analyzing portfolio behavior in diverse situations.	The quality of synthetic data heavily depends on the model used; if the model is inadequate, results may be distorted.
Synthetic data allow the generation of large datasets in a short time, accelerating the testing and improvement of strategies.	Synthetic data often lack the long-term historical context needed for assessing strategy resilience across different economic cycles.
Generative models enable control over volatility, correlations, and other key data parameters, creating idealized conditions for testing various hypotheses.	If a model generates data too similar to real conditions, it may lead to overfitting strategies to synthetic data, reducing their adaptability in real-world scenarios.

Based on the above, it can be concluded that generative models enable the simulation of rare market scenarios that are infrequent in historical data. This capability supports making informed investment decisions.

CONCLUSION

The use of synthetic data in forming investment portfolios under conditions of instability represents an approach that addresses the challenges of financial analysis. Generative algorithms create models of rare market scenarios that are inaccessible through the study of historical data. This method incorporates numerous factors shaping market dynamics, enabling the development of tools that account for current economic realities.

The application of synthetic data facilitates the creation of accurate models for risk assessment, allows for adjustments to asset management strategies under changing market conditions, and supports the development of solutions for addressing unconventional circumstances. Practical examples confirm the effectiveness of this approach in testing investment decisions aimed at reducing the likelihood of financial losses.

REFERENCES

1. Labiad B., Berrado A., Benabbu L. (2023), "Forecasting extreme events in the stock market using generative adversarial networks", International Journal of Intellectual Informatics Achievements, vol. 9, No. 2, pp. 218-230.
2. Naritomi Yu., Adachi T. (2020), "Increasing the volume of high-frequency financial data using a Generating Network", International Joint Conference on Web Intelligence and Intelligent Agent Technologies (WI-IAT), pp. 641-648.
3. D. Fernando, Stefan L., Boteanu B., Lamba S. And Ionescu. (2021), "Towards creating realistic financial time series using a generative link", 29th European Signal Processing Conference (EUSIPCO), pp. 1341-1345.
4. Dannels S. (2023), "Creating Catastrophes: Predicting recession using synthetic time series generated by GAN", arXiv, abs / 2302, 10490.
5. Tepelian R., Gopal A. (2023), "Generative machine learning for multidimensional equity returns", Proceedings of the Fourth ACM International Conference on AI in Finance, pp. 159-166.
6. Gibson L., Herger M., Kres D. (2023), "Generative flow-based model for modeling rare events", arXiv preprint arXiv:2305.07863.
7. Lee J. et al. (2020), "Generating realistic flows of stock market applications", Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, No. 01, pp. 727-734.
8. Colette. et al. (2021), "Towards realistic market modeling: an approach to generating interconnecting networks", Proceedings of the Second ACM International Conference on AI in Finance, pp.1-9.
9. Chen Yu. et al. (2021), "Towards the creation of synthetic multidimensional time series for forecasting spikes", Artificial Intelligence and Soft Computing: 20th International Conference, ICAISC 2021, Virtual Event, June 21-23, 2021, Proceedings, Part I, 20. – Springer International Publishing, pp. 296-307.
10. D. Fernando. et al. (2022), "Generation of realistic synthetic financial time series", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 18, No. 4, pp. 1-27.
11. Using the full capabilities of the AI generator in the banking sector[Electronic resource]. - Achievement mode: https://www.mckinsey.com/industries/financial-services/our-insights/capturing-the-full-value-of-generative-ai-in-banking.
12. Generative AI in banking: transformational potential and use cases that change the rules of the game [Electronic resource]. - Achievement mode: https://www.orientsoftware.com/blog/generative-AI-in-banking /.
13. Why is the generating AI in the banking sector

a secret resource: Your implementation plan [Electronic resource]. - Achievement mode: https://masterofcode.com/blog/generative-ai-in-banking.