

Synthetic Financial Time-Series Generation: A Comparative Study of Diffusion Models and GANs for Market Data Augmentation

Huzaifa Nasir¹ and Maaz Ali¹

Department of Computer Science, National University of Computer and Emerging Sciences
`{i221053, i221042}@nu.edu.pk`

Abstract. Financial markets generate complex time-series data with unique characteristics such as volatility clustering, fat-tailed distributions, and regime changes. Traditional machine learning models struggle with limited historical data, particularly for rare market events. This project proposes to investigate and compare modern generative models - specifically diffusion models and Generative Adversarial Networks (GANs) - for creating synthetic financial time-series data. We aim to evaluate their effectiveness in preserving statistical properties of real markets while enabling data augmentation for improved forecasting. Our approach includes comprehensive benchmarking against traditional methods including ARIMA, LSTM, Prophet, TimesFM, and TimesGPT, with evaluation using metrics such as MAPE and MAE. The project will employ STL decomposition for time-series analysis and assess practical utility through downstream financial applications.

Keywords: Financial Time-Series · Generative Models · Diffusion Models · GANs · Data Augmentation

1 Project Description

1.1 Problem Statement and Motivation

Financial time-series data is characterized by several unique properties that make it challenging for traditional machine learning approaches. These include non-stationarity, volatility clustering (where periods of high volatility are followed by high volatility), fat-tailed return distributions, and long-range dependencies [3][11]. Moreover, historical financial data is inherently limited, especially for rare but critical market events such as financial crises, flash crashes, or extreme volatility periods.

The scarcity of historical data for extreme market conditions poses significant challenges for risk management models, which need to accurately estimate tail risks and stress-test portfolios under various market scenarios. Traditional approaches like bootstrap sampling or parametric models (e.g., GARCH) often fail to capture the full complexity of market dynamics [6][1]. Recent advances in

financial machine learning [12] have highlighted the need for more sophisticated data generation techniques to address these limitations.

Recent advances in generative artificial intelligence, particularly Generative Adversarial Networks [8] and diffusion models [9], have shown remarkable capabilities in generating high-quality synthetic data across various domains. While these models were initially developed for image generation, they have been successfully adapted for time-series applications [17][14].

1.2 Research Objectives

This project aims to develop and evaluate a comprehensive framework for synthetic financial time-series generation with the following key objectives:

1. **Model Exploration:** Investigate and compare state-of-the-art generative models including diffusion models and GANs for financial time-series data synthesis, determining their effectiveness in capturing unique statistical properties of financial markets.
2. **Comprehensive Benchmarking:** Conduct thorough comparisons between generative approaches and traditional forecasting methods including ARIMA, LSTM, Prophet, TimesFM, and TimesGPT, evaluating their performance using metrics such as MAPE and MAE.
3. **Financial Data Analysis:** Employ STL decomposition and other time-series analysis techniques to understand the underlying patterns in financial data and assess how well different generative models preserve these characteristics.
4. **Practical Application:** Evaluate the utility of synthetic data in downstream financial tasks such as risk assessment, portfolio optimization, and forecasting, demonstrating real-world applicability.

1.3 Technical Approach

Our methodology will explore and compare multiple generative modeling paradigms for financial time-series synthesis:

Generative Model Investigation: We will investigate both diffusion models and Generative Adversarial Networks (GANs) as potential approaches for financial time-series generation. Diffusion models, such as Denoising Diffusion Probabilistic Models (DDPM), offer stable training and high-quality generation through iterative denoising processes [9][13]. GANs, particularly TimeGAN and similar architectures, provide efficient generation through adversarial training while maintaining temporal dependencies [17] [8].

Baseline Model Comparison: To establish comprehensive benchmarks, we will compare against various traditional and modern forecasting approaches including ARIMA models for classical time-series analysis [2], LSTM networks for deep learning baselines [10], Prophet for trend and seasonality modeling [15], TimesFM and TimesGPT for foundation model approaches [4] [7]. Additionally,

we will employ STL (Seasonal and Trend decomposition using Loess) for time-series decomposition analysis. Recent surveys on generative models for financial applications [16][5] and time-series generation [18] provide comprehensive background for our comparative analysis.

Evaluation Framework: The project will utilize multiple evaluation metrics including Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) for forecasting accuracy assessment, alongside statistical measures for distribution fidelity and financial stylized facts preservation. This comprehensive evaluation will guide the selection of the most suitable generative approach for financial data synthesis.

2 Main Functions and Expected Deliverables

2.1 Core System Functions

The proposed system will provide the following main functions:

1. Synthetic Data Generation

- Generate realistic financial time-series data (daily/intraday frequencies)
- Support multiple asset classes (equities, indices, cryptocurrencies)
- Maintain statistical properties of real financial data
- Enable flexible generation based on different model architectures

2. Comprehensive Model Evaluation

- Statistical fidelity assessment using distribution tests and moment matching
- Performance evaluation using MAPE, MAE, and other forecasting metrics
- STL decomposition analysis for trend and seasonality preservation
- Comparison with baseline models (ARIMA, LSTM, Prophet, TimesFM, TimesGPT)

3. Financial Data Analysis

- Time-series decomposition and pattern analysis
- Financial stylized facts verification (volatility clustering, fat tails)
- Regime detection and market condition analysis
- Cross-validation and robustness testing

4. Risk Management Applications

- Value-at-Risk (VaR) estimation using synthetic scenarios
- Stress testing with generated extreme market conditions
- Portfolio optimization with augmented datasets
- Backtesting enhancement through synthetic historical data

References

1. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* **31**(3), 307–327 (1986)
2. Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time series analysis: forecasting and control. John Wiley & Sons (2015)
3. Cont, R.: Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance* **1**(2), 223–236 (2001)
4. Das, A., Kong, W., Leach, A., Mathur, S., Sen, R., Yu, R.: A decoder-only foundation model for time-series forecasting. arXiv preprint arXiv:2310.10688 (2023)
5. Eckerli, F., Osterrieder, J.: Generative adversarial networks in finance: an overview. *Journal of Financial Data Science* **3**(2), 61–89 (2021)
6. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society* pp. 987–1007 (1982)
7. Garza, A., Mergenthaler-Canseco, M.: Timengpt-1. arXiv preprint arXiv:2310.03589 (2023)
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in neural information processing systems* **27** (2014)
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840–6851 (2020)
10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural computation* **9**(8), 1735–1780 (1997)
11. Mandelbrot, B.: The variation of certain speculative prices. *The journal of business* **36**(4), 394–419 (1963)
12. Prado, M.L.d.: Advances in financial machine learning. John Wiley & Sons (2018)
13. Rasul, K., Seward, C., Schuster, I., Vollgraf, R.: Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. *International Conference on Machine Learning* pp. 8857–8868 (2021)
14. Tashiro, Y., Song, J., Song, Y., Ermon, S.: Csd: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in Neural Information Processing Systems* **34**, 24804–24816 (2021)
15. Taylor, S.J., Letham, B.: Forecasting at scale. *The American Statistician* **72**(1), 37–45 (2018)
16. Wiese, M., Knobloch, R., Korn, R., Kretschmer, P.: Quant gans: deep generation of financial time series. *Quantitative Finance* **20**(9), 1419–1440 (2020)
17. Yoon, J., Jarrett, D., Van der Schaar, M.: Time-series generative adversarial networks. *Advances in neural information processing systems* **32** (2019)
18. Zhang, Y., Nie, F., Li, X.: A comprehensive survey on time series generative models. *ACM Computing Surveys* **55**(12), 1–37 (2023)