Synthetic Time Series Generation for Financial Applications: A Comparative Study of GANs, LLMs, and Econometric Models

Sha Fahad   Victor Sobottka   Alejandro Vacca

April 5, 2025

1 Project Description

This study explores the use of Generative AI for generating synthetic time series data to enhance model checking and back-testing in financial applications. The project will compare Generative Adversarial Networks (GANs), Large Language Models (LLMs), and traditional econometric approaches such as stationary bootstrap.

Given the increasing need for high-quality, privacy-preserving, cost-effective and shareable financial data, synthetic data generation has gained importance in AI-driven decision-making. This study aims to evaluate the effectiveness of different generative techniques in replicating real-world financial time series, using statistical and distributional distance metrics to measure likeness.

2 Key Objectives

- Implement and compare GAN-based, LLM-based, and econometric methods for synthetic time series generation.

- Develop quantitative metrics to assess the similarity between generated and real financial data.

- Explore practical applications, including data augmentation, model training, and privacy-preserving data sharing.

- Construct proof-of-concept implementations for financial applications.

- Compare the results to determine which one performs better

3 Importance

(Taken from YData AI)

In order to accelerate AI development and guarantee the best business practices and results, organizations rapidly need to become more data-centric. Quality and reliable data is the backbone of accurate decision-making but it can be expensive to collect or impractical to share, due to the privacy aspects associated.

Additionally, collecting real data can be costly due to acquisition, storage, and compliance expenses, making synthetic alternatives more viable.

For this reason, synthetic data has become a central topic with organizations: as it replicates the properties of real-world data, it can boost AI development in several ways, from data augmentation or model training and testing to data-sharing and privacy protection.

4 Related Work

- Using GANs for Sharing Networked Time Series Data: Challenges, Initial Promise, and Open Questions

- Generative Adversarial Nets for Synthetic Time Series Data

- YData Synthetic Data Documentation

- Time Series Forecasting with LLMs: Understanding and Enhancing Model Capabilities

- LLMs For Time Series Forecasting !!!

5 Considerations

- What is a GAN? What is a LLM?

- Compare GAN's generated synthetic time series against those generated by other "standard" econometric methods like stationary bootstrap.

- How do we compare? Distance distribution or other measures of likeness.

- Survey other models for generating synthetic time series, including Large Language Models (LLMs).

- Construct proof-of-concept models for generating synthetic time series to train models for financial applications (although some other domains may be considered such as climate, medical, or other).

6 Supervisor

This project is supervised by Argimiro Arratia.