ChatGPT

Using Diffusion Models in Generative AI for Financial Data Synthesis and Risk Management

Motivation and Context

Modern financial markets generate vast amounts of data, yet leveraging this data for risk analysis and scenario planning is challenging. Strict confidentiality and privacy regulations often limit the sharing of detailed financial datasets, hindering collaborative research and model development [1]. In addition, many financial institutions face a lack of sufficient high-quality data for training robust risk models, especially for rare or extreme events [2]. Generative AI has emerged as a promising approach to address these issues by creating synthetic financial data that preserves the statistical properties of real data while safeguarding privacy. This enables analysts to explore a broader range of market conditions and stress scenarios than what historical data alone can provide.

Challenges in Simulating Financial Data

Simulating realistic financial data is notoriously difficult due to the complex statistical patterns observed in markets. Asset prices and risk factors exhibit well-documented *stylized facts* – for example, asset returns have heavy tails (extreme fluctuations are more common than a normal distribution would predict) and volatility tends to cluster over time [3]. Such phenomena defy the simple assumptions of Gaussian distributions and independent increments. Furthermore, financial time series can show seasonality, regime shifts, and intricate cross-asset dependencies. Capturing these characteristics in simulations is crucial for credible risk management, yet even advanced generative models like GANs and VAEs have struggled to reproduce all such stylized facts simultaneously [3]. The result is that simple simulation methods often misrepresent tail risks or the timing of market turbulence, undermining their usefulness for stress testing and forecasting.

Limitations of Traditional Models

Traditional approaches for modeling and simulating financial data – including classical stochastic models and econometric techniques – come with significant limitations. Typically, one must assume a specific model (e.g. a Geometric Brownian Motion as in the Black–Scholes option pricing framework, or a GARCH volatility model for time-series data) and then calibrate its parameters to historical observations [4]. This **model-driven approach** imposes strong simplifying assumptions that often do not hold in reality: - **Simplified Dynamics:** For instance, the Black–Scholes model assumes constant volatility and lognormal price fluctuations, ignoring the fat tails and changing volatility observed in real markets [5]. Such assumptions can lead to mispricing of risk and an inability to foresee large swings. - **Calibration Challenges:** Models like GARCH (Generalized Autoregressive Conditional Heteroskedasticity) can capture time-varying volatility, but they still require choosing a functional form and rely on past data for calibration. If the historical sample is limited or the market undergoes structural changes, estimated parameters may be unreliable [6]. - **Limited Scope:** Many traditional models focus on a single asset or risk factor at a time, making it hard to simulate correlated multi-asset scenarios or truly novel stress events beyond the range of past observations. A standard Monte Carlo simulation with fixed covariances and drift, for example, will not generate unprecedented surprises outside historical experience [7].

Due to these limitations, conventional Monte Carlo simulations based on such models might fail to generate the full spectrum of possible outcomes, especially in the face of extreme or unforeseen events. This has prompted researchers and practitioners to seek more flexible, data-driven approaches.

Diffusion Models as a Generative Solution

Recent advances in generative AI offer a *data-driven* alternative to the above model-driven paradigm. In particular, **diffusion models** (a class of deep generative models originally developed for image synthesis) are now being applied to financial data generation with promising results. Unlike traditional approaches, diffusion models do not rely on explicit parametric assumptions about the data-generating process [8]; instead, they learn the distribution directly from historical time series, thereby capturing complex patterns and dependencies that fixed-form models might overlook. Diffusion models also tend to train more stably than earlier techniques like GANs, which often suffer from issues such as mode collapse. By iteratively adding and then removing noise during training, a diffusion model gradually learns to produce new samples that mimic the statistical properties of real financial data [9]. Early studies indicate that synthetic data generated in this way can replicate key distributional features of markets – including heavy-tailed returns and volatility clustering – with high fidelity [10]. In essence, diffusion models enable a "model-free" simulation approach: they let the data speak, which is particularly powerful in finance where the true data-generating process is complex and uncertain.

Project Scope and Objectives

Given this context, the present project focuses on leveraging diffusion-based generative models to create realistic synthetic financial data for risk management applications. The goal is to generate a wide range of plausible market scenarios that can bridge the gap between purely historical analyses and forward-looking scenario planning. In particular, the synthetic data and scenarios produced will support: - **Stress Testing and Scenario Analysis:** Expanding the set of stress scenarios beyond those seen in the past, to examine how portfolios or institutions might behave under extreme but plausible conditions [11] [12]. - **Hedging Strategy Evaluation:** Providing numerous simulated price paths to evaluate and verify hedging techniques, ensuring that risk mitigation strategies remain effective even under atypical market movements [11]. - **Data Augmentation for Modeling:** Generating additional datasets to train and validate risk models (e.g. for market risk or credit risk), which helps overcome data scarcity and improves the robustness of machine learning models [2].

Ultimately, synthesizing such diverse financial scenarios is expected to enhance risk assessment and decision-making processes by accounting for a broader range of uncertainties. This introduction has outlined the motivation and context for using generative AI in finance, the challenges of realistic financial data simulation, and the promise of diffusion models in addressing these issues. The subsequent background chapter will delve deeper into the relevant literature and technical methodologies – reviewing classical approaches and recent generative techniques, and explaining the theoretical foundations of diffusion models in the context of financial data generation.

---

[1] [9] [2309.01472] FinDiff: Diffusion Models for Financial Tabular Data Generation
https://arxiv.org/abs/2309.01472

[2] Harnessing Generative AI for Financial Services Innovation: What You Need to Know | Oscilar
https://oscilar.com/blog/generative-ai-financial-services

[3] Generation of synthetic financial time series by diffusion models
https://arxiv.org/html/2410.18897v1

4 6 8 uwaterloo.ca
https://uwaterloo.ca/computational-mathematics/sites/default/files/uploads/documents/shujie_liu_synthetic_asset_price_paths_generation_using_denoising_diffusion_probabilistic_model.pdf

5 Microsoft Word - 179814-fo
https://ibimapublishing.com/articles/JFSR/2018/179814/179814.pdf

7 11 How to Mine Synthetic Data: Pros and Cons of a Shiny New Tool for Risk Managers
https://www.garp.org/risk-intelligence/technology/how-to-mine-synthetic-data-pros-and-cons-of-a-shiny-new-tool-for-risk-managers

10 Beyond Monte Carlo: Harnessing Diffusion Models to Simulate Financial Market Dynamics by Andrew Lesniewski :: SSRN
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5039644

12 Generative artificial intelligence in finance (EN)
https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/12/generative-artificial-intelligence-in-finance_37bb17c6/ac7149cc-en.pdf