IMPERIAL

IMPERIAL COLLEGE LONDON

IMPERIAL-X

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

Author:
Simin Ali

Supervisor:
Dr Mikael Mieskolainen

Submitted in partial fulfillment of the requirements for the MSc degree in Artificial Intelligence Applications and Innovation of Imperial College London

September 2025

Contents

Chapter 1

Introduction

1.1 Motivation

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.

1.2 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.

1.3 Limitations of Traditional Models

Traditional approaches for modelling 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 [4].
- 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 [6].

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.

1.4 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 [4]; 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 [1].

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 [7]. 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.

Chapter 2

Background

Generative AI has emerged as a powerful tool in finance, enabling the creation of synthetic data that can augment scarce datasets and simulate complex market scenarios [1]. Recent advances in deep generative models, such as diffusion models, offer the ability to mimic the underlying distributions of financial data while preserving realistic statistical properties [1]. This literature review addresses the question: *How can diffusion-based generative models be effectively used for financial data synthesis and risk management, and what gaps remain?* To explore this, foundational theories of diffusion processes are examined, modern diffusion model architectures, applications in financial data generation and risk scenario analysis, and the methods used to evaluate synthetic data. The review synthesises findings from high-quality sources including peer-reviewed journals and top conferences (e.g. NeurIPS, ICML), aiming to identify key methodologies and open research challenges. By critically assessing prior work, it is outlined how the proposed research will build on these insights to advance the use of diffusion models for financial risk analysis.

2.1 Theoretical Foundations

Diffusion models, specifically *Denoising Diffusion Probabilistic Models* (DDPMs), are generative models grounded in stochastic processes [8]. They employ a two-phase Markov chain: a forward diffusion process that gradually perturbs data with noise, and a reverse diffusion process that learns to recover data from noise [8].

The forward process is often defined as an Ornstein–Uhlenbeck (OU) process—a continuous-time diffusion where data X_t is steadily driven toward random noise. A common formulation is an OU process with drift $-\frac{1}{2}X_t$ and Gaussian increments [9]. In the limit as $t \to T$, this process yields a simple known distribution, typically a standard Gaussian.

The reverse-time process is described by a stochastic differential equation (SDE) that uses the score function $nablalogp_t(x)$ the gradient of the data log-density—to incrementally denoise the sample [9]. In practice, one trains a neural network to estimate this score (or equivalently, to predict the noise added) at each timestep. Generating a sample entails simulating the learned reverse SDE, starting from random noise and iteratively refining it into a synthetic data point [9].

This approach draws on concepts from non-equilibrium thermodynamics and stochastic calculus, linking modern diffusion models to classical ideas such as Brownian motion and Langevin dynamics [9].
Stochastic processes also underlie traditional financial models. For instance, the Black–Scholes model assumes asset prices follow a Brownian motion with drift. However, empirical research shows that real financial time series exhibit fat-tailed distributions and volatility clustering that deviate from Brownian motion assumptions[3]. These persistent statistical patterns (dubbed *stylized facts*) include heavy-tailed return distributions (power-law tails), autocorrelated volatility (periods of high volatility clustering together), and other effects like periodic seasonality [3]. Capturing such complex behaviour is challenging for simple diffusions or parametric models. Diffusion models in AI leverage stochastic noise but with trainable reverse dynamics, giving them the expressive power to model these non-Gaussian, path-dependent characteristics. The theoretical foundation of diffusion generative models thus provides a probabilistic framework well-suited to mirror the randomness in financial systems. By basing generation on SDEs, these models naturally incorporate notions of randomness and time-evolution akin to those in quantitative finance, bridging the gap between machine learning and traditional stochastic finance theory.

2.2 Architectures and Training Methods

Modern diffusion models pair the above stochastic framework with advanced neural network architectures. A common architecture is a U-Net convolutional network (originally popularized in image segmentation) used as the *denoising model* at each timestep. For example, image-based diffusion models like DDPM train a U-Net to take a noisy image and time-step as input and output a denoised image estimate [3]. This architecture, often enhanced with multi-head attention, enables capturing both local and global structure during generation. In time-series contexts, researchers adapt the architecture: some use 1-D convolution or Transformer encoders to handle sequential data. An example is TimeGrad, which applies an autoregressive diffusion for multivariate time-series forecasting by iteratively generating one step at a time [10]. In TimeGrad, the model learns to transform white noise into a sample of the next return distribution, leveraging a variational bound on likelihood to learn gradients (scores) of the data distribution [10]. This autoregressive strategy breaks a high-dimensional problem (forecasting an entire path) into a series of 1-step diffusion generations, improving tractability for financial time series.
Training methods for diffusion models differ significantly from GANs or VAEs. Diffusion models are trained by *denoising score matching* or equivalently minimizing a reweighted variational loss across diffusion timesteps [10]. In practice, the training objective is often to predict the Gaussian noise added to a sample at a given time t; the loss is the mean squared error between the predicted noise and true noise, averaged over random t. Notably, this yields stable training dynamics – there is no adversarial min–max game as in GANs, which mitigates mode collapse and improves coverage of the data distribution [3]. Indeed, recent studies found that diffusion models surpass GANs and VAEs in generating high-quality *and* diverse out-

puts [3]. This superior fidelity and diversity is attributed to the diffusion training process which encourages modelling of all modes of the data (including tail events) rather than collapsing to the most probable mode.
A key architectural extension is conditional diffusion modelling, which incorporates auxiliary information y (such as class labels or, in finance, market conditions) to guide generation [9]. Conditioned diffusion models modify the network to also accept y as input (e.g. through concatenation or a cross-attention mechanism) so that the reverse process samples from $p(x|y)$. In the financial domain, conditioning enables directed scenario generation – for instance, generating price paths under a given volatility regime. *CoFinDiff* (Controllable Financial Diffusion) demonstrates this by using cross-attention to condition on summary statistics like trend and realized volatility [11]. By encoding these conditions and attending to them during denoising, the model learns to produce samples that conform to specified market scenarios [11]. This architectural innovation addresses a key limitation of earlier generative models, which often lacked the ability to steer simulations toward user-defined scenarios [11]. Finally, while diffusion models achieve strong results, a noted drawback is sampling speed – generation requires hundreds or thousands of iterative denoising steps, which is *time-consuming* compared to one-shot GAN sampling [9]. Research into *accelerated samplers* (e.g. by skipping or merging diffusion steps) is ongoing [9], and future architectural improvements may focus on efficiency to make diffusion models more practical for real-time risk management.

2.3 Financial Data Synthesis

Synthetic financial data generation has been attempted with various generative models over the past decade. Early approaches used Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) to learn the distribution of historical financial time series and sample new plausible sequences [3]. These efforts aimed to reproduce the stylized facts of markets (fat tails, volatility clustering, etc.) in simulated data [3]. For example, Wiese *et al.* (2020) and Eckerli & Osterrieder (2021) applied GAN-based models to create realistic price series, while Brophy *et al.* (2023) explored VAEs for similar purposes [3]. Despite some success, none of these earlier models could capture *all* key stylized facts simultaneously [3]. Generating data that obeys every statistical nuance of real markets remained an open challenge – GANs often struggled with mode collapse or unstable training, and hand-crafted VAE latent spaces could miss extreme events.
Diffusion models have recently been proposed as an alternative approach to financial data synthesis due to their stability and expressiveness [3]. Takahashi and Mizuno (2024) introduced a diffusion-based method to generate synthetic financial time series that exhibit multiple stylized facts [3]. Their approach uniquely integrates wavelet transforms: multiple time series (e.g. stock prices, trading volumes, and bid-ask spreads) are transformed into a joint time-frequency representation (akin to an image) on which a DDPM is trained [3]. Because diffusion models excel at image generation, converting time series into spectrogram-like images allowed the model to learn complex temporal patterns via 2D convolutional networks [3]. After gener-

ation, an inverse wavelet transform converts the images back into time series. This method enabled *simultaneous modelling of correlated series* (price, volume, spread) by treating them as different color channels of an image[3]. The authors report that the diffusion model accurately reproduces stylized facts in unison across these series (e.g. heavy-tailed price returns with matching volatility clustering in volumes)[3], marking a notable advance in realism of synthetic financial data.

For tabular financial data (e.g. credit portfolios, balance sheets), researchers have likewise applied diffusion models. *FinDiff* (Sattarov *et al.*, 2023) is a diffusion model designed for mixed-type tabular data generationin financial contexts [1]. FinDiff uses embedding layers to represent categorical features alongside numerical ones, then diffuses and denoises in this joint feature space [1]. A key insight from FinDiff is handling entity-specific data: Liu *et al.* (2024) point out that many financial tables have an ”entity” column (such as a firm or customer ID) which previous models treated as just another category, neglecting its special role [12]. They propose *EntTabDiff*, which explicitly learns a distribution over entities and then conditions the diffusion model on a chosen entity to generate entity-specific records [12]. This two-step approach (generate synthetic entities, then generate data conditioned on them) preserves inter-entity heterogeneity – a critical factor when synthesizing, say, bank customer data or multi-company financial statements. Both FinDiff and EntTabDiff demonstrate that diffusion models can handle complex *mixed data* (numerical, categorical, multi-entity) common in finance, achieving high fidelity to real data distributions while maintaining privacy[1].

Notably, diffusion models have also been applied to augment scarce financial datasets. Because regulatory or proprietary data are often limited, synthetic data can help train robust models. For instance, *CoFinDiff*(Tanaka *et al.*, 2025) addresses data scarcity in rare but high-risk events (e.g. market crashes) by allowing conditional generation of such scenarios[11]. By learning with conditions (trend, volatility) and leveraging the reversibility of wavelet-transformed sequences, CoFinDiff can produce diverse samples of extreme events that were too few in the historical data[11]. These synthetic samples can then be used to stress test algorithms or train trading strategies under conditions that are plausible yet under-represented in the original data. Across these works, a pattern emerges: diffusion models are expanding the frontier of financial data synthesis, yielding more realistic and controllable synthetic data than previously possible. This lays a foundation for improved risk analysis, as discussed next.

2.4 Applications of Gen AI to Risk Management and Scenario Generation

One primary application of generative models in finance is the creation of market scenarios for risk management. Banks and financial institutions conduct stress tests and risk forecasting by simulating how portfolios behave under various hypothetical market conditions. Generative AI enhances this by learning complex joint distributions of risk factors from historical data and sampling new *coherent* scenarios that go beyond simple parametric assumptions. Early work by Flaig and Junike (2022) showed that a GAN-based economic scenario generator could replicate the outputs of a full market risk model over a one-year horizon, providing a data-driven alternative to regulatory risk modelling approaches[13]. Diffusion models now offer further improvements in scenario generation due to their ability to capture tail risks and fine-grained dependencies. For example, *CoFinDiff* allows controllable scenario generation – users can specify a severe downtrend with high volatility, and the model will produce synthetic price paths aligning with that condition [11]. This is particularly useful for stress testing: risk managers can generate many variants of extreme but plausible scenarios (such as flash crashes or liquidity crises) and evaluate portfolio resilience under each. Tanaka *et al.* (2025) demonstrate that CoFinDiff-generated data not only capture known stylized facts of markets, but also improve the training of risk management models (like deep hedging strategies) by exposing them to a richer set of adverse scenarios [11].

Synthetic data generated by diffusion models can be used for model validation: risk teams can test their predictive models on artificially generated worst-case scenarios to see if they hold up, a practice recommended to complement standard backtesting [14].

Key applications of generative AI to financial risk management include:

- Economic Scenario Generation and Stress Testing: Producing multi-factor market scenarios for what-if analysis (e.g. joint simulations of stock prices, interest rates, and volatilities) under extreme conditions[1]. Diffusion models like FinDiff can generate scenarios for regulatory stress tests while respecting realistic cross-asset correlations and distributions[1].

- Data Augmentation for Risk Models: Augmenting limited datasets (e.g. for credit risk or operational risk) with additional synthetic examples. By training on larger and more diverse data – including rare events generated by models – risk prediction models can become more robust to market shifts [14]. This also aids in model validation, allowing tests of risk models on scenario data covering conditions not seen in the original sample.

- Privacy-Preserving Data Sharing: When multiple institutions need to collaborate (such as central banks aggregating risk data), diffusion models can generate anonymous synthetic datasets that preserve statistical properties but protect confidential information[1]. For instance, regulators can share synthetic bank data for systemic risk analysis without revealing any single institution's records.

Collectively, these applications illustrate that generative AI – and diffusion models in particular – are becoming integral in risk scenario generation and management. They provide flexible tools to explore "what-if" questions, improve risk metric calculations, and ensure models are resilient to the full spectrum of market behaviours, including those not witnessed before. The next consideration is how to rigorously evaluate these synthetic outputs for reliability and realism.

2.5 Evaluation Metrics and Validation

Evaluating the quality and utility of synthetic financial data is critical before it can be trusted for risk management. Researchers employ a variety of metrics and validation approaches to assess generative models:

- Fidelity: This refers to how closely the synthetic data distribution matches the real data distribution. It can be evaluated through statistical tests and distance measures. Common checks include comparing moments (mean, variance, skewness), marginal and joint distributions, and verifying the presence of stylized facts (e.g. does synthetic data exhibit fat tails and volatility clustering similar to real data?[3]. In Takahashi & Mizuno's work, for example, they explicitly show that synthetic series obey power-law tails and long-memory volatility dynamics, confirming high fidelity to real market behavior[3].

- Diversity (Coverage): A good generative model should produce a wide variety of outcomes, not just repeating a few patterns. This is often measured by assessing if the synthetic dataset covers the support of the real data without mode collapse. Metrics like the Maximum Mean Discrepancy or clustering-based techniques can quantify coverage. In diffusion model evaluations, *diversity* is sometimes evidenced by better coverage of extreme events or minor modes compared to GANs[11]. Chen *et al.* (2024) note that diffusion models allow *versatile guidance* without sacrificing diversity, which is a key advantage in controlled scenario generation. Forest-Diffusion (a tabular diffusion model) was explicitly evaluated on diversity, alongside fidelity and other criteria [12].

- Utility for Machine Learning: This measures how useful the synthetic data is for downstream tasks. For instance, one can train a predictive model on synthetic data and test it on real data (or vice versa) to see if performance holds up. A high *utility* score means the synthetic data effectively supplements or replaces real data for model training[12]. FinDiff's results showed that models trained on its synthetic data achieved comparable accuracy on real data tasks, indicating the synthetic data retained practical information[1]. Similarly, CoFinDiff reported that hedging models trained on its enriched scenarios outperformed those trained on real data alone[11]. Such evidence speaks to the *validity* of the synthetic data for risk modelling purposes.

- Privacy and Regulatory Compliance: Especially in finance, it's crucial that synthetic data does not leak sensitive information about real entities. Metrics for privacy include measures like attribute disclosure risk or membership inference tests. A properly tuned diffusion model should generate samples that are statistically similar to real data *in aggregate* but not identifiable as any specific real record. FinDiff, designed with regulator use-cases, emphasizes privacy alongside fidelity and utility[1]. In evaluations, synthetic records are checked to ensure no real transaction or customer can be reconstructed from them (e.g., using distance to nearest real data point as a heuristic).

- Domain-Specific Validation: Beyond generic metrics, domain-informed tests are applied. For risk management, this may involve backtesting synthetic-data-driven models. For example, if synthetic scenarios are used to estimate VaR, one can test how often actual losses exceed those VaR predictions (the Kupiec test) to ensure the synthetic approach is neither too optimistic nor overly conservative. Scenario plausibility checks by subject matter experts are also valuable – generated scenarios should make economic sense (no arbitrage violations or implausible correlations). Some studies perform stress outcome analysis, verifying that known historical crises are within the support of generated scenarios (if a model never reproduces something like the 2008 crash in many simulations, it might be underestimating tail risk).

In summary, a combination of statistical, practical, and expert-driven evaluation is used to validate synthetic financial data. High-quality studies report multiple such metrics: for instance, a recent survey noted that tabular diffusion models are judged on *fidelity, diversity, machine learning utility, and statistical similarity* to original data [12]. When diffusion models manage to score well on all these fronts, it increases confidence that the synthetic data is both *realistic* and *useful*. Nevertheless, rigorous validation remains an ongoing requirement. Especially for risk management, any synthetic data or generative model-driven decision must undergo thorough testing to ensure it does not introduce unseen biases or understate risks.

2.6 Research Gap

The literature surveyed demonstrates that diffusion models have rapidly become a leading approach for generative AI in finance, owing to their strong theoretical basis and empirically observed performance. Foundational work bridged stochastic diffusion processes with deep learning, providing models that can capture the rich statistical structure of financial data more faithfully than earlier GAN or VAE-based methods. State-of-the-art diffusion architectures, often leveraging innovative techniques like wavelet-based image conversion or cross-attention conditioning, have pushed synthetic data generation to new heights, achieving realistic replication of stylized facts and enabling controllable scenario outputs. Applications in risk management are already evident: diffusion models contribute to improved risk measure estimation, scenario analysis, and data augmentation in ways that can enhance the resiliency of financial models[11]. These advances come with rigorous evaluation protocols that assess fidelity and diversity which reflects a maturing field that recognizes the need for trustworthy synthetic data[12][1].
Despite this progress, the literature also highlights several gaps and challenges. One major challenge is the controllability vs. complexity trade-off: early generative models lacked scenario control, a gap addressed by works like CoFinDiff, but there remains room to expand the range of conditions under which generation can be directed (e.g. conditioning on macroeconomic indicators or textual news events is still in nascent stages). Another issue is computational efficiency – diffusion models are computationally intensive, and while methods to accelerate sampling exist[9],

integrating these models into real-time risk monitoring systems will require further innovation in model distillation or step-reduction techniques. Moreover, evaluation standards need continued development. For instance, while studies evaluate stylized facts and use generic metrics, there is no universal benchmark for financial synthetic data quality and this research will aim to contribute a more standardised evaluation framework, possibly drawing on domain-specific risk metrics and regulatory guidelines.

The proposed research will build upon these insights to contribute novel value in two ways. First, it is important to address the gap in controllability by developing a diffusion-based framework that can incorporate richer conditional information (such as *multi-factor economic scenarios or stress narratives*) directly into the generation process. By doing so, a tool can be created that risk managers can query for bespoke scenario generation – an advancement over current models limited to a few technical condition inputs. Second, the focus will be on integrating diffusion models with existing risk management workflows, examining how synthetic data impacts risk forecasts and portfolio stress tests when used in conjunction with traditional models. This involves not only generating data but also providing guidelines for its prudent use, validation, and interpretation. This work aims to ensure that the power of diffusion generative models can be harnessed safely and effectively in the financial sector. By addressing current gaps, such as improving conditional generation fidelity, speeding up sampling, and formalising validation, this research will push the frontier of how generative AI can contribute to financial data science and risk management. Ultimately, this contributes to a more robust understanding of risk by enabling analyses that were previously infeasible due to data limitations, thus offering novel insights and practical tools for both academics and practitioners in quantitative finance.

Chapter 3

Project Plan

3.1 Timeline

The remainder of the project is divided into four phases. Phases I–IV cover data preparation, model development, evaluation, and defence preparation.

Phase I: Toy Experiments and Small-Scale Benchmarks (June 5–June 28)

- **June 5–June 14:**
  - Implement minimal "toy" versions of each core method (e.g., a small DDPM on a 1D Gaussian mixture, a simple GAN on a toy distribution, and a tabular diffusion on a tiny dataset).
  - Verify each toy model visually (e.g., compare learned vs. true density) and compute a basic metric (e.g., KS distance).
- **June 15–June 23:**
  - Collect a small slice of real returns (e.g., 2,000 data points), split into training and test sets.
  - Train baseline models (GARCH, toy-GAN) on the slice; generate synthetic test-set returns.
  - Evaluate and compare using a few key metrics: KS distance, ACF of squared returns, and a simple VaR backtest.
- **June 24–June 28:**
  - Set up core infrastructure (GPU environment, code repository structure).
  - Finalize high-level design for the full diffusion model (discrete DDPM vs. score-based SDE, network backbone, noise schedule).
  - Consolidate evaluation scripts for subsequent phases.

Phase II: Full Diffusion Model Development (June 29–July 26)

- **Late June – Early July:**
  - Implement the full diffusion pipeline (forward noise corruption, reverse denoising network).
  - Integrate data loader and ensure correct input/output shapes.
  - Train a small-scale prototype on a subset of historical returns to confirm functionality.
- **Mid–Late July:**
  - Train the diffusion model on the entire training dataset (e.g., all available historical returns) using a full noise schedule.
  - Save checkpoints and monitor convergence (denoising loss).
  - Generate preliminary synthetic paths for initial evaluation.

Phase III: Refinement and Comprehensive Evaluation (July 27–August 30)

- **Late July – Early August:**
  - Refine noise schedule or network architecture to better capture tails and volatility clustering.
- **Mid–Late August:**
  - Generate large batches of synthetic trajectories and compute detailed metrics:
    - Distributional fidelity (KS, MMD, tail indices).
    - Volatility autocorrelations (ACF of squared returns).
    - Risk measures (VaR, ES) with backtesting (Kupiec/Christoffersen tests).
  - Perform stress-scenario sampling, compare extreme loss quantiles to historical events.
  - Document all findings in tables and plots for the Results and Discussion chapters.

Phase IV: Thesis Write-Up and Defence Preparation (August 31–September 17)

- **August 31–September 6:**
  - Integrate all chapters into the full thesis document.
- **September 7–September 11:**

  - Submit draft to advisor, incorporate feedback.
  - Finalise code repositories.

- **September 12–September 17:**
  - Prepare and rehearse defence presentation.
  - Refine answers to anticipated questions and finalise slides.

- **September 18:**
  - Thesis defence.

3.2 Success Criteria

3.2.1 Deliverable-Related Success

1. Background Report submitted by June 4, 2025 (Chapters 1–4 complete, formatted, and proofread).

2. Toy Experiments completed by June 14, 2025, with documented results verifying that each method recovers its respective toy distribution.

3. Small-Slice Benchmarks completed by June 28, 2025 (GARCH and toy GAN comparisons on a 2,000-point slice).

4. Diffusion Model Prototype (mini training on 1D toy and small slice) functioning by July 5, 2025.

5. Full Diffusion Model Training finished by July 26, 2025 (converged on full historical data).

6. Comprehensive Evaluation (Phase III) completed by August 30, 2025, including all quantitative tables and figures for Results chapter.

7. Full Thesis submitted for advisor review by September 11, 2025.

8. Successful Defence on September 18, 2025.

3.2.2 Academic Contributions

1. Reproducible Code - All toy experiments, small-slice implementations, and full diffusion code must be uploaded to a public GitHub repository with clear instructions by August 31, 2025.

2. Comparative Analysis Write-Up - A thorough, reproducible comparison of diffusion vs. GARCH and GAN, with tables and plots showing each metric, completed in the Results chapter (due August 30, 2025).

3. Demonstration of Controllable Scenario Generation (if conditioning implemented) - At least one use case where specifying a target volatility or stress label yields a synthetic ensemble that meets that stress specification within $\pm 10\%$ of the target statistic.

4. Successful Thesis defence - Convincing presentation of methodology, results, and limitations on September 18, 2025.

Bibliography

[1] Timur Sattarov, Marco Schreyer, and Damian Borth. FinDiff: Diffusion Models for Financial Tabular Data Generation, September 2023. URL http://arxiv.org/abs/2309.01472. arXiv:2309.01472 [cs]. pages 1, 2, 4, 7, 8, 9, 10

[2] Saurabh Bajaj. Harnessing Generative AI for Financial Services Innovation: What You Need to KnowOscilar, October 2023. URL https://oscilar.com/blog/generative-ai-financial-services. pages 1

[3] Tomonori Takahashi and Takayuki Mizuno. Generation of synthetic financial time series by diffusion models, October 2024. URL https://arxiv.org/html/2410.18897v1. pages 1, 5, 6, 7, 9

[4] Shujie Liu. Synthetic Asset Price Paths Generation Using Denoising Diffusion Probabilistic Model. pages 2

[5] Zuzana Janková. Drawbacks and Limitations of Black-Scholes Model for Options Pricing. *Journal of Financial Studies and Research*, 2018:1–7, August 2018. ISSN 2166000X, 2166000X. doi: 10.5171/2018.79814. URL https://ibimapublishing.com/articles/JFSR/2018/179814/. pages 2

[6] How to Mine Synthetic Data: Pros and Cons of a Shiny New Tool for Risk Managers, . URL https://www.garp.org/risk-intelligence/technology/how-to-mine-synthetic-data-pros-and-cons-of-a-shiny-new-tool-for-risk-managers pages 2

[7] Andrew Lesniewski. Beyond Monte Carlo: Harnessing Diffusion Models to Simulate Financial Market Dynamics, 2024. URL https://www.ssrn.com/abstract=5039644. pages 3

[8] Harnessing Generative AI for Financial Services Innovation: What You Need to KnowOscilar, . URL https://oscilar.com/blog/generative-ai-financial-services. pages 4

[9] Minshuo Chen, Song Mei, Jianqing Fan, and Mengdi Wang. Opportunities and challenges of diffusion models for generative AI. *National Science Review*, 11 (12):nwae348, November 2024. ISSN 2095-5138, 2053-714X. doi: 10.1093/nsr/nwae348. URL https://academic.oup.com/nsr/article/doi/10.1093/nsr/nwae348/7810289. pages 4, 5, 6, 10

[10] Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8857–8868. PMLR, July 2021. URL `https://proceedings.mlr.press/v139/rasul21a.html`. ISSN: 2640-3498. pages 5

[11] Yuki Tanaka, Ryuji Hashimoto, Takehiro Takayanagi, Zhe Piao, Yuri Murayama, and Kiyoshi Izumi. CoFinDiff: Controllable Financial Diffusion Model for Time Series Generation, March 2025. URL `http://arxiv.org/abs/2503.04164`. arXiv:2503.04164 [q-fin]. pages 6, 7, 8, 9, 10

[12] Zhong Li, Qi Huang, Lincen Yang, Jiayang Shi, Zhao Yang, Niki van Stein, Thomas Bäck, and Matthijs van Leeuwen. Diffusion Models for Tabular Data: Challenges, Current Progress, and Future Directions, February 2025. URL `http://arxiv.org/abs/2502.17119`. arXiv:2502.17119 [cs] version: 1. pages 7, 9, 10

[13] Solveig Flaig and Gero Junike. Scenario generation for market risk models using generative neural networks, August 2023. URL `http://arxiv.org/abs/2109.10072`. arXiv:2109.10072 [cs]. pages 8

[14] Ishita Kaur. Generative AI for Financial Risk Management Models, August 2024. URL `https://www.crossml.com/generative-ai-for-financial-risk-management-models/`. Section: Blogs. pages 8