Deep Generation of Financial Time Series

From GANs to Diffusion Models for Private Credit

PhD Seminar in Quantitative Finance

January 2026

Learning Objectives

By the end of this seminar, you will be able to:

1. Understand stylized facts of financial time series and why they matter for generative modeling
2. Master the architectures of GANs, VAEs, and Diffusion models
3. Apply deep generative models to private credit data generation
4. Evaluate synthetic data quality using appropriate metrics
5. Implement VaR estimation and stress testing with synthetic data

Prerequisites:

- Deep learning fundamentals (neural networks, backpropagation)
- Probability theory (distributions, expectations, KL divergence)
- Time series analysis (stationarity, autocorrelation)

Target audience: PhD students in finance, economics, or quantitative fields

Why Synthetic Financial Data?

Key Challenges

- Privacy: GDPR, data protection regulations
- Scarcity: Limited historical data for rare events
- Opacity: Private markets lack transparency

Applications

- Data augmentation for ML models
- Stress testing and scenario analysis
- Model validation and backtesting
- Privacy-preserving data sharing

Market Context

- Private credit: \$1.5T+ AUM globally
- Growing demand for alternative data
- Regulatory pressure for model validation

Research Impact

- 3,000+ papers on generative finance (OpenAlex)
- Rapid growth since 2019 (TimeGAN)
- CFA Institute 2025 report on synthetic data

Synthetic data enables research and applications impossible with real data alone

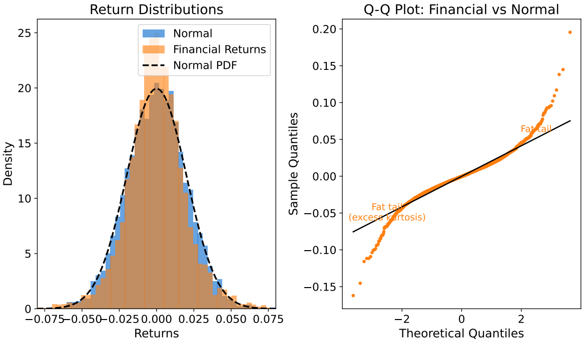
Financial Time Series: Stylized Facts

Key Properties

- Fat tails (excess kurtosis > 3)
- Volatility clustering
- Leverage effect
- Absence of linear autocorrelation
- Slow decay of ACF in $|r_t|$

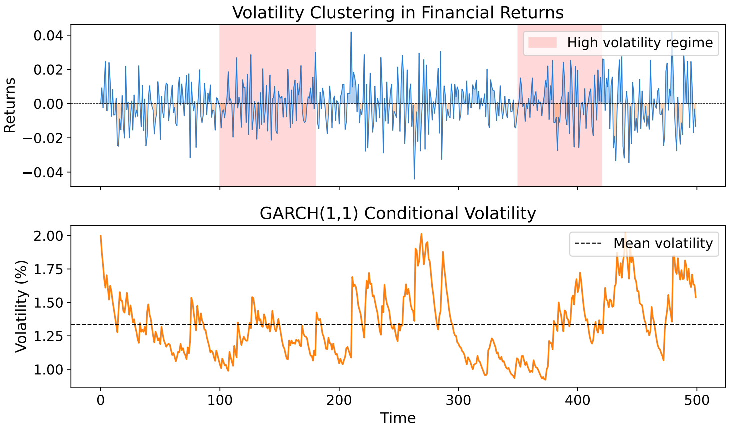
Challenge for Generative Models

- No model captures ALL facts
- Trade-off: quality vs diversity



A good generative model must reproduce these empirical regularities

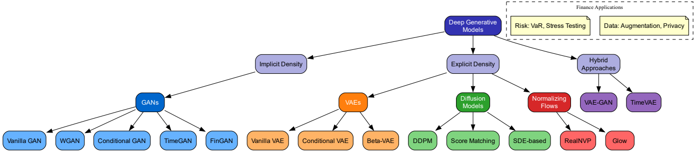
Fat Tails and Volatility Clustering



GARCH(1,1) Model: $\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$

Volatility clustering: large changes tend to be followed by large changes

Deep Generative Model Taxonomy



Finance Applications

Risk: VaR, Stress Testing

Data: Augmentation, Privacy

Deep Generative Models

Implicit Density

Explicit Density

Hybrid Approaches

GANs

VAEs

Diffusion Models

Normalizing Flows

VAE-GAN

TimeVAE

Vanilla GAN

WGAN

Conditional GAN

TimeGAN

FinGAN

Vanilla VAE

Conditional VAE

Beta-VAE

DDPM

Score Matching

SDE-based

RealNVP

Glow

Implicit density (GANs) vs Explicit density (VAEs, Diffusion, Flows)

Generative Adversarial Networks: Architecture

Components

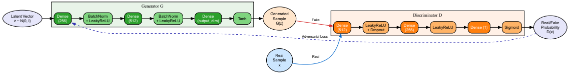
- Generator G: $z \to x_{fake}$
- Discriminator D: $x \to [0, 1]$

Minimax Objective

$$\min_G \max_D \mathbb{E}_x[\log D(x)] + \mathbb{E}_z[\log(1 - D(G(z)))]$$

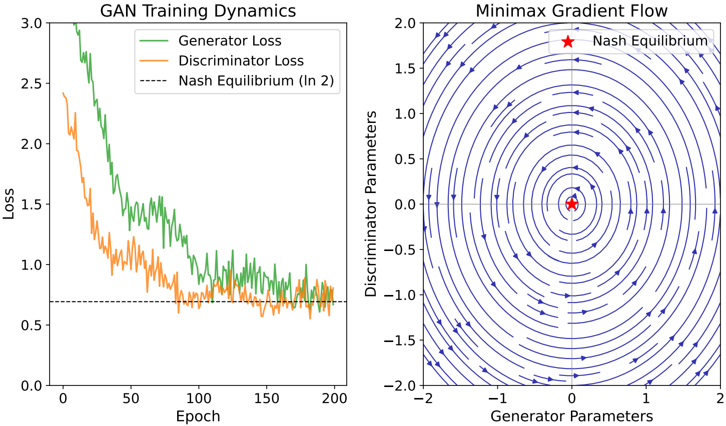
Training

- Alternate between D and G updates
- Nash equilibrium: $D(x) = 0.5$



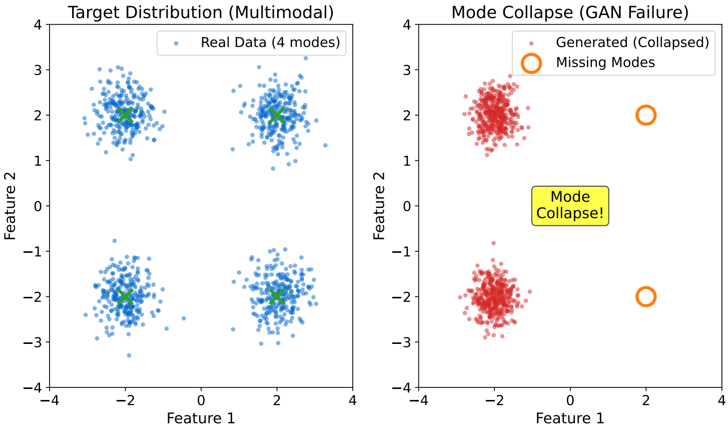
GANs learn to generate by fooling a discriminator

GAN Training Dynamics



At Nash equilibrium, D cannot distinguish real from fake: $D(x) = 0.5$

Mode Collapse: A Critical GAN Failure Mode



Solutions: WGAN, spectral normalization, minibatch discrimination

Mode collapse: generator produces limited variety, missing parts of data distribution

Wasserstein GAN with Gradient Penalty

Wasserstein Distance

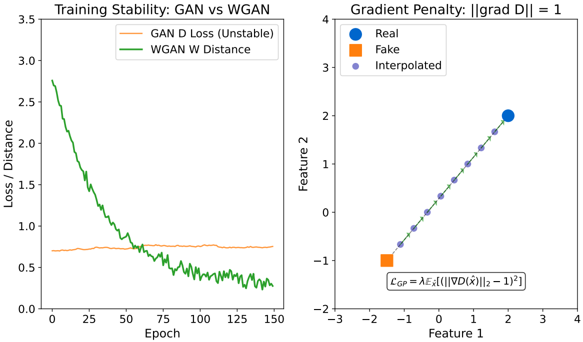
$$W(P_r, P_g) = \sup_{||f||_L \leq 1} \mathbb{E}_{x \sim P_r}[f(x)] - \mathbb{E}_{x \sim P_g}[f(x)]$$

Gradient Penalty

$$\mathcal{L}_{GP} = \lambda \mathbb{E}_{\hat{x}}[(||\nabla_{\hat{x}} D(\hat{x})||_2 - 1)^2]$$

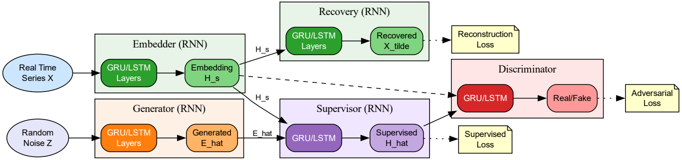
Benefits

- Stable training
- Meaningful loss metric
- No mode collapse



WGAN-GP enforces 1-Lipschitz constraint via gradient penalty

TimeGAN: Temporal Generative Adversarial Network

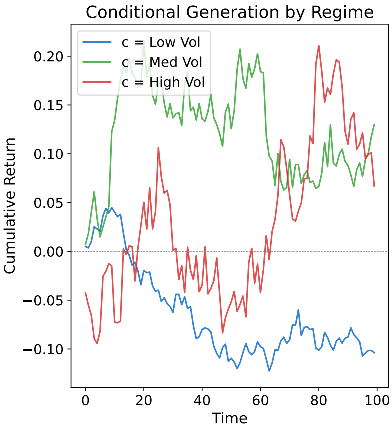


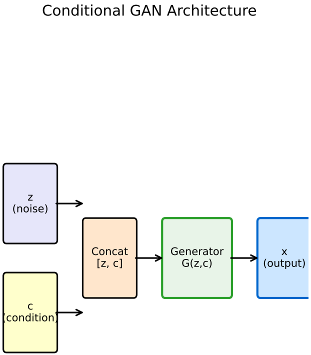
Key Innovation: Combines RNN-based autoencoding with adversarial training for temporal data

---

TimeGAN (Yoon et al., 2019): State-of-the-art for time series generation

Conditional GANs: Regime-Specific Generation





Objective: $\min_G \max_D \mathbb{E}[\log D(x|c)] + \mathbb{E}[\log(1 - D(G(z|c)|c))]$

Condition on volatility regime, sector, or other attributes

Variational Autoencoder: Architecture

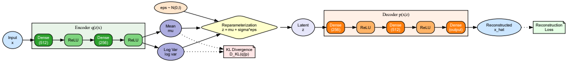
Components

- Encoder $q_\phi(z|x)$: inference network
- Decoder $p_\theta(x|z)$: generative network
- Prior $p(z) = \mathcal{N}(0, I)$

Reparameterization Trick

$$z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$$

Enables gradient flow through sampling



VAEs provide explicit density estimation and smooth latent space

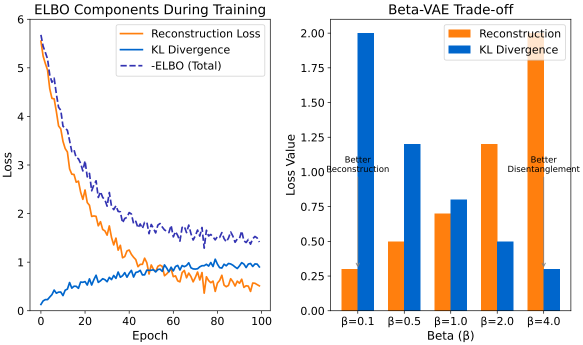
Evidence Lower Bound (ELBO)

ELBO Decomposition

$$\mathcal{L} = \underbrace{\mathbb{E}_q[\log p_\theta(x|z)]}_{\text{Reconstruction}} - \underbrace{D_{KL}(q_\phi(z|x)||p(z))}_{\text{Regularization}}$$

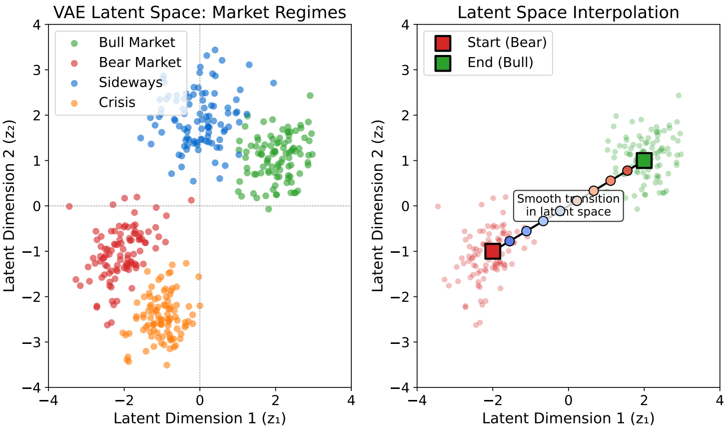
Trade-off

- Reconstruction: fit the data
- KL: regularize latent space
- β-VAE: $\mathcal{L} = \text{Recon} - \beta \cdot \text{KL}$



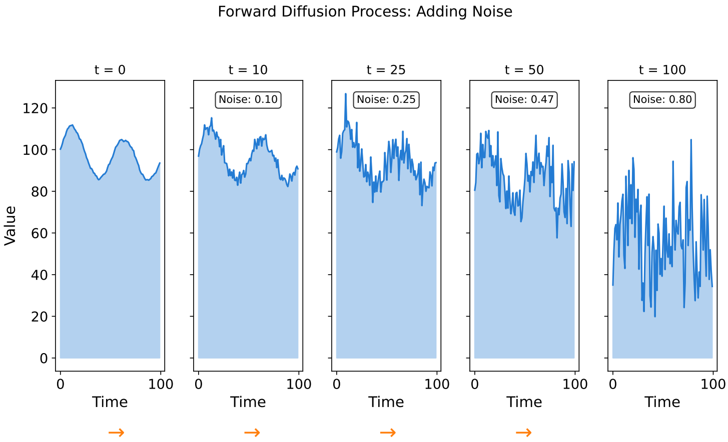
ELBO maximization = maximizing log-likelihood lower bound

Latent Space: Market Regime Encoding



Smooth latent space enables interpolation between market regimes

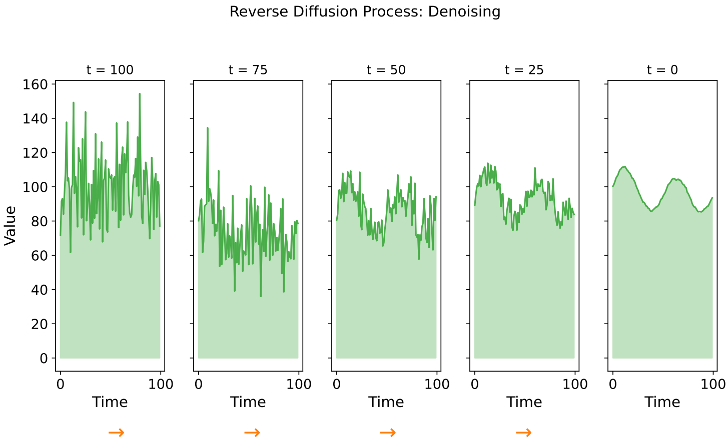
Diffusion Models: Forward Process



Forward: $q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I)$

Gradually add noise until data becomes pure Gaussian noise

Diffusion Models: Reverse Process



Reverse: Learn to denoise via neural network $\epsilon_\theta(x_t, t)$

Generate samples by iteratively denoising from pure noise

Score-Based SDE Formulation

Forward SDE

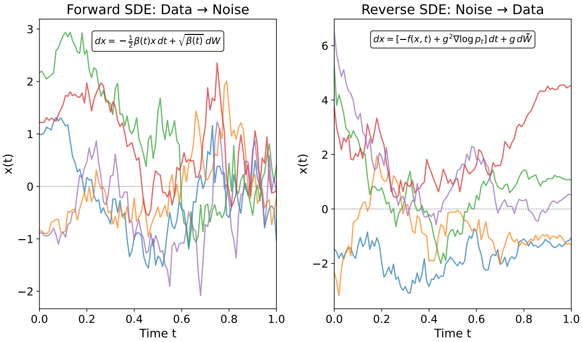
$$dx = f(x,t)dt + g(t)dW$$

Reverse SDE

$$dx = [f(x,t) - g^2(t)\nabla_x \log p_t(x)]dt + g(t)d\bar{W}$$

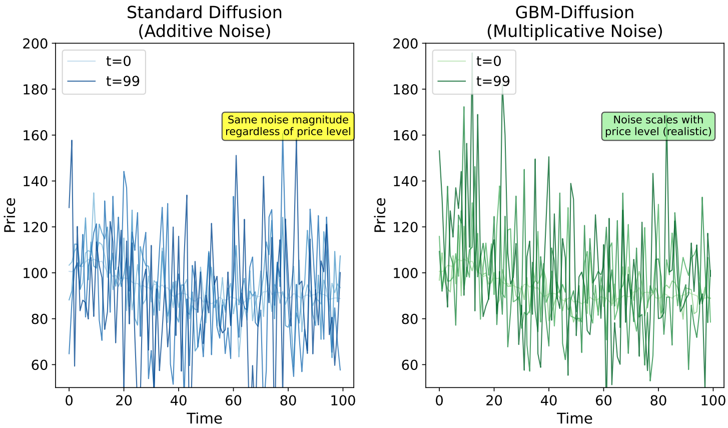
Score Function

$$s_\theta(x,t) \approx \nabla_x \log p_t(x)$$



Score matching learns the gradient of the log-density

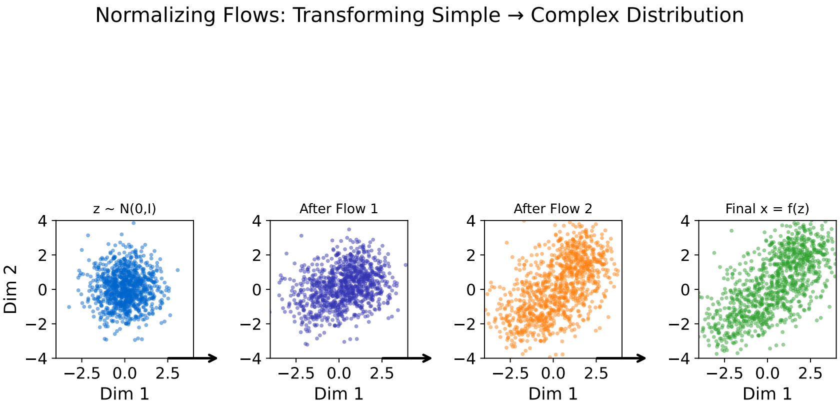
GBM-Diffusion: Finance-Specific Adaptation



Key Insight: Inject noise proportionally to price level (heteroskedasticity)

GBM-diffusion better reproduces financial stylized facts (QF 2025)

Normalizing Flows: Invertible Transformations



Change of Variables: $p_X(x) = p_Z(f^{-1}(x)) \cdot |\det \frac{\partial f^{-1}}{\partial x}|$

Flows transform simple distributions into complex ones via invertible maps

Private Credit: Data Challenges

Market Characteristics

- Illiquid, infrequent transactions
- Quarterly NAV reporting (lag)
- Limited historical data
- Heterogeneous deal structures

Data Types

- Loan-level: spreads, defaults, recoveries
- Fund-level: NAV, cash flows, IRR

Why Synthetic Data?

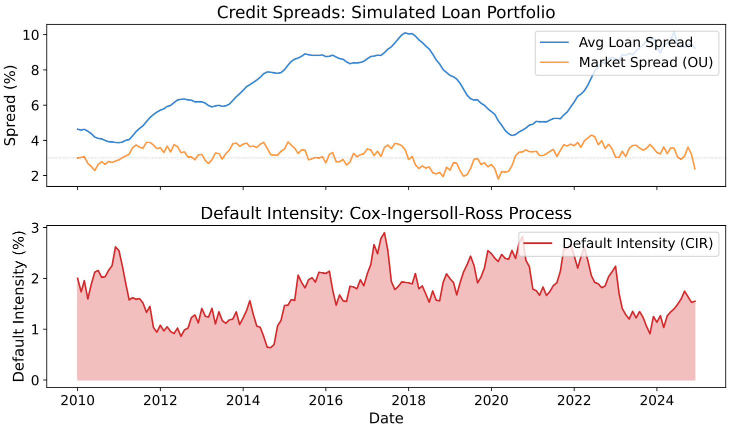
- Augment sparse historical records
- Generate stress scenarios
- Privacy-preserving sharing
- Model training and validation

Simulation Approach

- Stochastic processes (OU, CIR)
- Copula-based dependencies
- J-curve modeling for funds

Private credit AUM exceeds $1.5T globally with limited data transparency

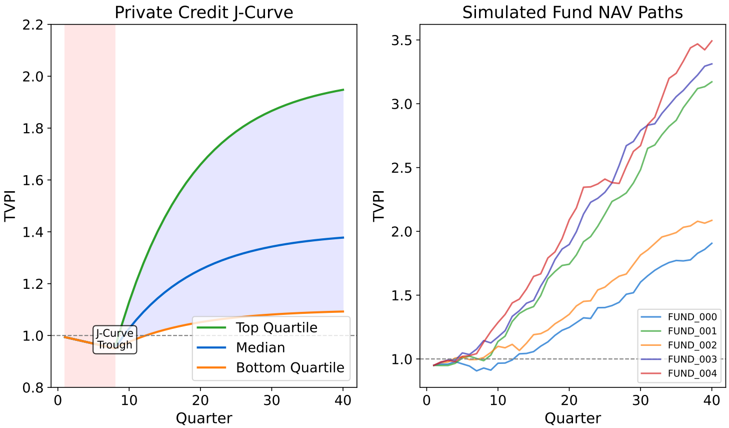
Loan-Level Simulation: Spreads and Defaults



Models: Ornstein-Uhlenbeck (spreads), Cox-Ingersoll-Ross (default intensity)

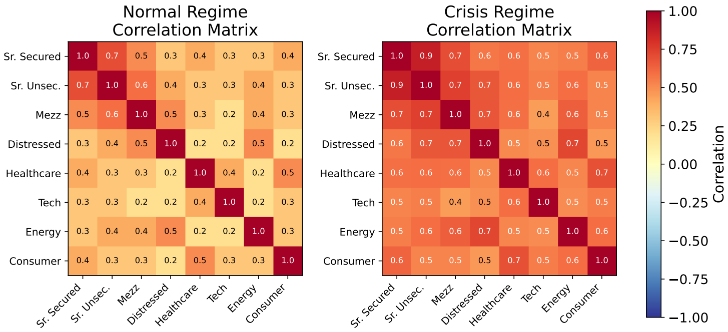
Mean-reverting processes capture realistic credit dynamics

Fund-Level Simulation: J-Curve and NAV



J-curve: initial value decline followed by recovery as investments mature

Correlation Structure: Normal vs Crisis Regimes



Normal Regime Correlation Matrix

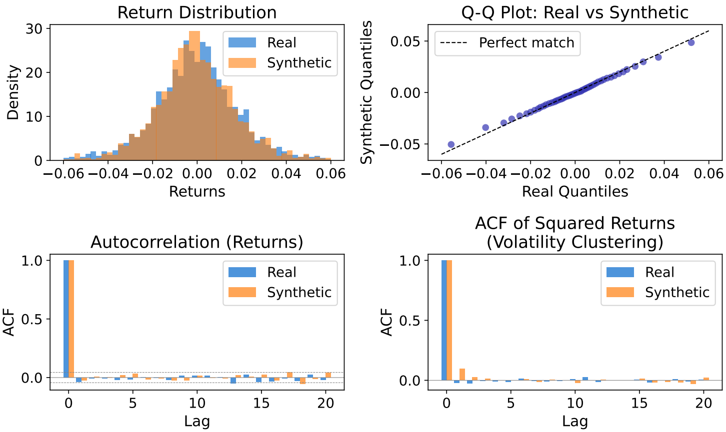
	Sr. Secured	Sr. Unsec.	Mezz	Distressed	Healthcare	Tech	Energy	Consumer
Sr. Secured	1.0	0.7	0.5	0.3	0.4	0.3	0.3	0.4
Sr. Unsec.	0.7	1.0	0.6	0.4	0.3	0.3	0.4	0.3
Mezz	0.5	0.6	1.0	0.5	0.3	0.2	0.4	0.3
Distressed	0.3	0.4	0.5	1.0	0.2	0.2	0.5	0.2
Healthcare	0.4	0.3	0.3	0.2	1.0	0.4	0.2	0.5
Tech	0.3	0.3	0.2	0.2	0.4	1.0	0.2	0.3
Energy	0.3	0.4	0.4	0.5	0.2	0.2	1.0	0.3
Consumer	0.4	0.3	0.3	0.2	0.5	0.3	0.3	1.0

Crisis Regime Correlation Matrix

	Sr. Secured	Sr. Unsec.	Mezz	Distressed	Healthcare	Tech	Energy	Consumer
Sr. Secured	1.0	0.9	0.7	0.6	0.6	0.5	0.5	0.6
Sr. Unsec.	0.9	1.0	0.7	0.7	0.6	0.5	0.6	0.5
Mezz	0.7	0.7	1.0	0.7	0.6	0.4	0.6	0.5
Distressed	0.6	0.7	0.7	1.0	0.5	0.5	0.7	0.5
Healthcare	0.6	0.6	0.6	0.5	1.0	0.6	0.5	0.7
Tech	0.5	0.5	0.4	0.5	0.6	1.0	0.5	0.5
Energy	0.5	0.6	0.6	0.7	0.5	0.5	1.0	0.6
Consumer	0.6	0.5	0.5	0.5	0.7	0.5	0.6	1.0

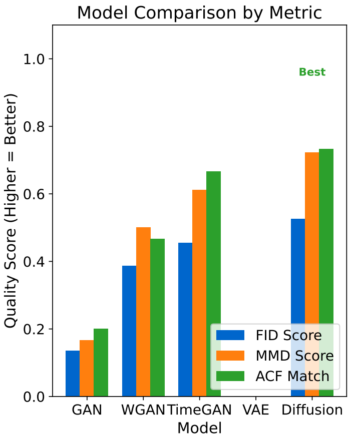
Crisis regime: correlations increase (diversification breakdown)

Synthetic vs Real: Distribution Matching



Good synthetic data matches marginal distributions AND temporal dependencies

Evaluation Metrics for Synthetic Data



EVALUATION METRICS FOR SYNTHETIC DATA

1. FID (Frechet Inception Distance)
    - Measures distance between real and synthetic feature distributions
    - Lower = Better
    - Captures both quality and diversity

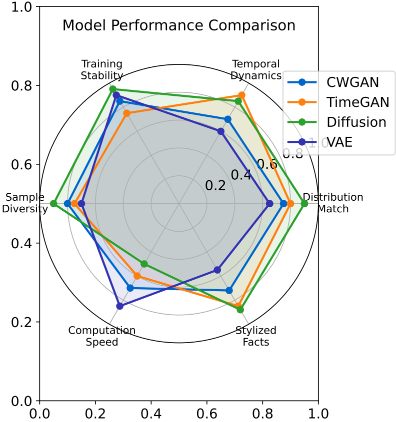
2. MMD (Maximum Mean Discrepancy)
    - Kernel-based distribution comparison
    - Lower = Better
    - Non-parametric, flexible

3. ACF Test (Autocorrelation Function)
    - Compares temporal dependencies
    - Lower error = Better
    - Critical for time series

4. Additional Metrics:
    - Stylized Facts Tests
    - Discriminative Score
    - Predictive Score
    - Privacy Metrics

Multiple metrics needed: no single metric captures all quality aspects

Model Comparison: Which Architecture to Choose?

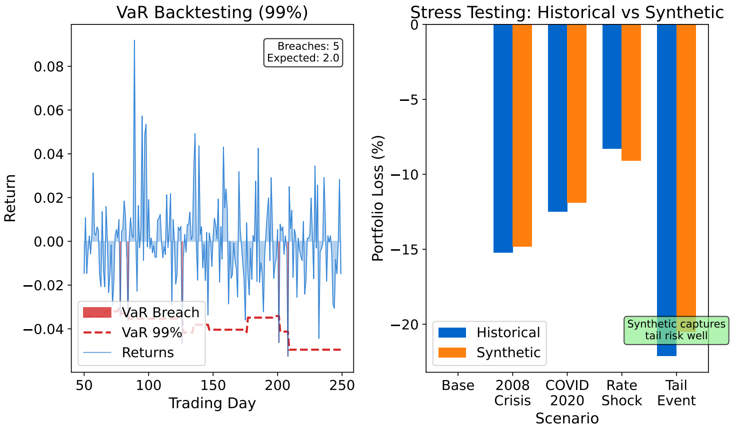


Model Summary

Model	Best For	Weakness	Speed
CWGAN	Conditional gen.	Mode collapse	Fast
TimeGAN	Temporal deps.	Complex training	Medium
Diffusion	Quality & diversity	Slow sampling	Slow
VAE	Latent space	Blurry outputs	Fast

Diffusion models lead in quality; TimeGAN best for temporal dynamics

Application: VaR and Stress Testing



Synthetic data enables robust tail risk estimation and scenario analysis

Research Frontiers and Open Problems

Unsolved Challenges

- No model satisfies ALL stylized facts
- Privacy preservation metrics underdeveloped
- Evaluation standards not established
- Mode collapse in complex distributions

Emerging Directions

- Conditional diffusion with cross-attention
- Foundation models for finance
- Federated learning with synthetic data

Application Frontiers

- Real-time synthetic data generation
- Causal structure preservation
- Multi-asset joint generation
- Regulatory sandbox testing

Key Papers (2024-2025)

- QF 2025: GBM-Diffusion
- arXiv 2401.10370: VaR Review
- CFA Institute: Synthetic Data Report

Active research area with significant opportunities for PhD contributions

Conclusion and Key Takeaways

Summary

1. Deep generative models (GANs, VAEs, Diffusion) can generate realistic financial time series
2. Each architecture has trade-offs: quality vs speed vs stability
3. Private credit benefits from synthetic data for augmentation and stress testing
4. Evaluation requires multiple metrics: FID, MMD, ACF tests, stylized facts

Recommended Reading

- Yoon et al. (2019): TimeGAN
- Wiese et al. (2020): Quant GANs
- Quantitative Finance 2025: GBM-Diffusion
- CFA Institute 2025: Synthetic Data in Investment Management

Repository: `github.com/Digital-AI-Finance/deep-generation-of-financial-time-series`

Questions? Visit the GitHub repository for discussion