Executive Summary

Context and motivation

Financial institutions increasingly rely on stress testing to evaluate portfolio resilience under extreme but plausible market conditions. Since the 2008 Global Financial Crisis, supervisory frameworks such as Basel III and the European Banking Authority's guidelines have institutionalized these exercises, emphasizing scenario realism, multi factor consistency, and tail-risk awareness.

Despite regulatory progress, the analytical foundations of stress testing remain rooted in two outdated paradigms: historical replay and parametric dependance models. Historical scenarios reproduce past cries directly, like the 2008 or 2020 market conditions, but fail to anticipate novel configurations of systemic risk. Parametric models, including Gaussian copulas and multivariate normal factor frameworks, impose rigid correlation structures that underestimate tail co-movements and collapse under regime shifts. These methods can only explore variations in what markets have already experienced, leaving regulators and risk managers exposed to the unknown structural combinations of risk that have yet to materialize.

Advances in deep generative modeling offer a data-driven alternative. Models such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and their time-series variants can learn complex joint distributions directly from historical data, capturing nonlinear and dynamic dependencies. By generating synthetic but statistically coherent trajectories, these models can expand the space of plausible market scenarios. The thesis explores how such models, specifically TimeGAN, can be adapted for empirical financial stress testing.

Research Aim and Objectives

The primary aim of this thesis is to evaluate whether deep generative models can produce synthetic financial time series that are both empirically realistic and practically useful for portfolio-level stress testing.

More concretely, the research pursues the following objectives:

a) Replicate multivariate market dynamics beyond what traditional replay or copula-based approaches can capture.
b) Incorporate macro-volatility regime information to stabilize training and improve realism.
c) Asses tail-risk metrics, such as VaR and CVaR, in synthetic data compared with real historical behavior.
d) Evaluate predictive and discriminative performance, establishing whether TimeGAN can generalize to unseen market configurations.

Through these objectives, the study addresses four research questions and tests three hypotheses concerning empirical realism, stability, and tail-risk accuracy.

The empirical framework is based on a multi-asset dataset encompassing equities, bonds, commodities, and currencies, complemented by macro-risk factors such as VIX, MOVE, and CDX indices. The final aligned dataset spans from 2004-2025, containing 4,408 rolling windows of 60-day sequences and 26 standardized features.

Rigorous preprocessing ensured analytical integrity: Missing values were left unfilled in asset prices (to preserve empirical realism) but filled in risk factors via forward-backward interpolation. Log-returns were computed to normalize volatility across instruments. A union trading calendar was applied across all markets. A macro-volatility regime label was created using 60-day rolling clustering on risk factors, categorizing data into low, medium, and high volatility states.

Methodological Framework

This study implements TimeGAN, a hybrid model combining recurrent autoencoders and adversarial learning. The model architecture comprises five key components: Embedder and recovery networks for latent representation learning, Generator and Supervisor for temporal sequence synthesis, and Discriminator for adversarial realism enforcement.

Hyperparameters were tuned for stability and empirical alignment: λ_adv = 5.0, λ_sup = 1.0, λ_rec = 5.0, λ_gp = 1.0 and a generator discriminator update ratio of 2:1.

Training proceeded in three phases: embedding pretraining, supervised sequence modeling, and joint adversarial refinement, executed in Google Colab using the Conda-managed environment deepgen.

Subsequent refinements introduced variance widening, latent noise injection, and direct generator outputs (bypassing the recovery step), all designed to mitigate mode collapse and over-smoothing commonly observed in early GAN applications to finance.

Results and Observations

Quantitative evaluation focused on both distributional and temporal realism. Across progressive model versions, the following empirical patterns emerge:

- Distributional alignment: The final model versions achieved a mean Kolmogorov-Smirnov (KS) statistic of 0.28 across features, indicating strong overlap between real and synthetic marginal distributions.
- Temporal coherence: The lag-1 autocorrelation gap ($|\Delta ACF|$) dropped from ≈ 1.0 in early collapsed runs to ≈ 0.14, demonstrating realistic volatility clustering.
- Predictive fidelity: A Mean Squared Error (MSE) of ≈ 0.0169 in one-step-ahead forecasting confirmed that synthetic sequences preserved realistic short-term predictability.

- Tail-risk comparability: Synthetic portfolio metrics produced VaR and CVaR within ± 10% of historical benchmarks, with similar drawdown frequency distributions.

Visual diagnostics such as t-SNE projections (see figure 5) showed that synthetic sequences gradually expanded from collapsed manifolds into rich, overlapping clusters resembling real data structures. Collectively, these results validate the three hypotheses:

I. TimeGAN generated data is indistinguishable from rea1 data (H1).
II. Regime conditioning improves stability and realism (H2).
III. The framework enhances tail-risk estimation accuracy relative to historical replay (H3)

Implications for Risk Management

The findings demonstrate that deep generative models can augment conventional stress testing by generating coherent, data-driven scenarios that reflect modern market complexities. For financial institutions, this means the ability to simulate forward-looking crises beyond past experience, evaluate portfolio sensitivity to nonlinear cross-asset shocks, and design stress scenarios dynamically adjusted to volatility regimes.

Regulators and risk managers can integrate such generative frameworks to complement, rather than replace existing methodologies, bridging the gap between empirical realism and policy interpretability. The generated data also holds promise for Monte Carlo stress testing, liquidity simulations, and tail-risk calibration in capital planning.