Latent Representations of Financial Data for Market Regime Analysis

14 February 2026

Periods of financial market stress are rare, which leads to a scarcity of historical data that represent extreme conditions. This leads to arising of challenges for derivative pricing, risk management, and stress testing, where realistic stressed scenarios are required. A major quantitative feature that describes market expectations is the implied volatility surface (IVS). Such a surface $\sigma_{\text{impl}}(m, \tau)$, depends on time to maturity τ and log-moneyness $m = \log(K/S_0)$, where K is a strike price and S_0 initial value of the underlying asset. The IVS reflects market perception of risk. Generating economically consistent synthetic IVS data under stress is therefore important from a practical point of view.

Recent work in [2] proposes a variational autoencoder (VAE) framework to address data scarcity. The resulting latent space provides a low-dimensional representation of IVS dynamics, allowing controllable generation of economically interpretable surfaces. Empirical results demonstrate that the model captures stylized features characteristic of stressed regimes.

Building on this framework, the central research question of this thesis is whether one can detect and characterize the boundary between "normal" and "stressed" market regimes in the latent space. In particular, we aim to determine whether regime transitions correspond to geometric structures such as mode boundaries or singular sets.

To address this question, the thesis will employ the AE-OT methodology introduced in [1]. In this framework, latent codes $\{z_i\}_{i=1}^N$ are interpreted as discrete target points of a semi-discrete optimal transport (OT) problem. Given a source distribution μ (e.g., a continuous reference measure in latent space), the OT map is characterized by a convex potential whose cells form a power diagram. Each cell corresponds to a mode of the data distribution, and boundaries between cells are found via transport potentials $\langle z_i, x\rangle + h_i$ for suitable height parameters h_i. These boundaries constitute lower-dimensional singular sets. In generative modeling, such structures provide a mechanism to detect mode separation and avoid mode collapse or mixing.

The objective of this thesis is therefore:

- to develop a geometrically grounded methodology for detecting boundaries between latent modes in implied volatility surfaces data.
- to analyze the geometric structure of the latent space learned by VAE.
- to explore relevant applications, for example, investigating whether the resulting OT cell partitions and singular sets correspond to economically meaningful distinctions between normal and stressed market conditions. If such distinction can be found, generate new synthetic samples for stressed market periods.

Student:	Anita Daliga
Daily supervisor:	Shuaiqiang Liu
Responsible supervisor:	Martin van Gijzen

References

[1] Dongsheng An, Yang Guo, Na Lei, Zhongxuan Luo, Shing-Tung Yau, and Xianfeng Gu. AE-OT: A new generative model based on extended semi-discrete optimal transport. In *International Conference on Learning Representations*, 2020.

[2] Jing Wang, Shuaiqiang Liu, and Cornelis Vuik. Controllable generation of implied volatility surfaces with variational autoencoders, 2025.