

PhD Progress Report

Gabin Taibi

Behavioural, Management and Social sciences (BMS), University of Twente

Promotor(s): prof. dr. Joerg Osterrieder

Co-Promotor(s): dr. Xiaohong Huang

Supervisors: dr. Stefan Schlamp, dr. Axel Gross-Klussmann

November 8, 2025

Thesis Title: *Modeling Narrative Dynamics for Volatility Regime Detection in Financial Markets*

PhD Period: December 2023 – November 2027

1. Introduction

My PhD investigates how financial narratives influence volatility and market regime changes. I study how information contained in unstructured text (news headlines, articles, speech transcripts, corporate communication or filings) can be quantified and linked to volatility patterns. The objective is to integrate natural language processing, econometrics, and machine learning to model narrative formation, diffusion, and their relationship with market uncertainty. Ultimately, the goal is to quantify the informational role of narratives in shaping risk perception and to assess their predictive power for structural breaks and volatility regimes in financial markets. The project combines theoretical and empirical approaches, aiming to provide a framework that bridges narrative economics and quantitative finance.

2. Research Progress and Planning

I have completed the first stage of my PhD, which focused on data collection and the systematic literature review. I implemented custom libraries to collect and structure datasets from RavenPack (news headlines), LSEG (earnings call transcripts), BIS (worldwide central bank speeches), SEC EDGAR (10-K and 10-Q filings), and Deutsche Börse (nanosecond-level Xetra and Eurex trading data). I developed Python pipelines for data ingestion, preprocessing, and version control. The first paper, an AI-enhanced systematic literature review on financial narratives, is currently under revision at *Financial Innovation*.

The second stage, which is ongoing, focuses on the detection and quantification of financial narratives. I am developing both supervised and unsupervised NLP frameworks using embedding similarity, Large Language Models (LLM), and graph-based clustering algorithms. This part of the research will lead to two papers, one presenting the narrative detection framework and another introducing **TOPol**, a tool used for semantic polarity shift detection.

The third stage, starting in 2026, will consist in computing and benchmarking realized volatility estimators using Deutsche Börse data. The **realized-library** I created provides efficient Python and C++ implementations for realized volatility estimators and jump detection. I will also compute historical implied volatility, and realized and implied roughness and vol-of-vol. This work will produce two papers: one on the impact of high-frequency traders on market quality and another on volatility estimation methods (eventually a third one on Hurst exponent computation).

The fourth stage will integrate narratives into volatility modeling. I will apply structural break detection methods such as Bai–Perron, CUSUM, and Bayesian frameworks to realized and implied volatility measures. The goal is to link potential regime shifts with narrative dynamics changes and to test for causal relationships. The final stage of the PhD will synthesize these results and present a comprehensive understanding of narrative-driven volatility dynamics.

Envisioned Chapters:

- **Chapter 1 – Systematic Literature Review:** This chapter defines and contextualizes the concept of financial narratives. It traces how narrative ideas have evolved within economics and finance, highlighting the shift from qualitative, discourse-based interpretations toward quantitative, data-driven modeling.
- **Chapter 2 – Narrative Detection and Quantification:** This chapter develops a robust methodological framework for identifying, quantifying, and tracking financial narratives across diverse text sources. It applies supervised approaches, such as LLM or similarity-based tagging of predefined economic and financial themes, alongside unsupervised topic modeling and large language model-based discovery of emerging narratives. The analysis covers several levels of granularity: market-wide narratives derived (from media), macroeconomic narratives (from regulatory communications), and firm-specific narratives (from corporate filings). The outcome is a set of dynamic narrative intensity indices capturing how the prominence and sentiment of specific narratives evolve over time.
- **Chapter 3 – Volatility Modeling and Benchmarking:** This chapter builds the empirical foundation for volatility analysis using high-frequency trading data. It constructs a comprehensive volatility modeling pipeline incorporating realized and implied volatility, as well as roughness measures. The framework compares various estimators such as realized variance, MinRV, and MedRV or kernel-based, while also estimating volatility-of-volatility and Hurst exponents to assess persistence in volatility dynamics.
- **Chapter 4 – Narrative–Volatility Dynamics:** The final empirical chapter integrates the textual and market-based components developed in previous chapters. It investigates whether fluctuations in narrative intensity, sentiment, and thematic composition correspond to structural changes in volatility regimes. Using econometric and machine learning techniques, the chapter tests causal and predictive relationships between narrative and volatility components. Structural break detection frameworks are also applied to examine whether narratives act as leading indicators of volatility transitions.

3. Achievements and Activities

During the first half of my PhD, I developed several research projects and tools. The `realized-library` provides advanced realized volatility estimators for high-frequency data, while `TOPol` models semantic and polarity drift in textual narratives. I also built an AI-enhanced tool for systematic literature discovery and synthesis using the OpenAlex API and vector-based clustering.

My research includes strong industry collaborations. At Deutsche Börse, I conducted a research internship focusing on the classification of high-frequency traders and their impact on market quality. This work forms the basis of my first empirical paper. I am also collaborating with Quoniam Asset Management on the use of narrative-based alpha signals derived from news headlines, which will result in a joint research publication.

I have one paper under revision and several others in progress, including a preprint of `TOPol` already available on arXiv and a study on high-frequency market microstructure. I also actively engaged in the COST Action 19130 on Fintech and AI in Finance (including event coordination, budget planning, network communication, etc.) and the MSCA Doctoral Network on Digital Finance. I also participated in PhD training schools and in conferences on various topics such as finance, AI, and financial econometrics, and will have completed more than half of the required 30 ECTS before Q1 2026.