PhD Qualifier Report

Gabin Taibi

Faculty, Department: Behavioural, Management and Social sciences (BMS), High-tech Business and Entrepreneurship (HBE)
Promotor(s): prof. dr. Joerg Osterrieder
Co-Promotor(s): dr. Xiaohong Huang
Supervisors: dr. Stefan Schlamp, dr. Axel Gross-Klussmann
Qualifier Committee Members: prof. dr. Wolfgang Haerdle, prof. dr. Ali Hirsa, prof. dr. Daniel Pele

November 10, 2025

Table of Contents

Abstract

Modeling Narrative Dynamics for Volatility Regime Detection in Financial Markets

- **Research question:** How do forms of narratives influence volatility and regime changes in markets?
- **Research summary:** Financial markets are increasingly driven by narratives, defined as collective interpretations that influence expectations, volatility, and market regimes. Yet, despite their central role, narratives remain largely unquantified in financial modeling. This research aims to bridge that gap by developing a computational framework linking the evolution of financial narratives to volatility dynamics and structural market shifts. The thesis integrates methods from natural language processing, high–frequency econometrics, and machine learning to (1) detect and quantify narratives across multiple textual sources, (2) measure volatility and its higher–order properties from high–frequency data, and (3) analyze the causal and predictive relationship between narrative shifts and volatility regimes.
- **PhD expected period:** Dec 1st, 2023 - Nov 30th, 2027

Research Motivation

Rationale:

- Narrative information forms a latent context for market decisions (Shiller), yet quantitative models often overlook these textual dynamics.
- Recent advances in transformer-based NLP:
  - Allow scalable narrative extraction, extending analysis beyond simple sentiment/polarity.
  - As LLMs increase textual data, this research seeks to identify which narratives matter most.
- Access to high-quality, high-frequency (HFT) data allows for precise estimation of market dynamics, including volatility and its higher-order properties.
- Understanding how narrative shocks translate into volatility regime changes provides academic and practical insights for risk management and narrative-aware trading.

Research Motivation

Social perspectives:

- 1. Understanding Financial Narratives and Financial Markets Reactions;
- 2. Improving Financial Stability and Crisis Prediction;
- 3. Empowering Retail and Institutional Decision-Making;
- 4. Interdisciplinary Applications Beyond Finance.

Research Motivation

Scientific perspective:

- 1. Improving NLP methods for financial text;
- 2. Advancing the quantitative modeling of various forms of narratives within finance;
- 3. Highlight interconnexion between narratives or forms of narratives (e.g. how macro and micro information spread in news);
- 4. Linking narrative and volatility dynamics.

PhD Thesis Plan Overview

Stage	Chapter	Theme	Duration
I	Introduction & Chapter 1	Data Collection, Literature Review, and NLP–Finance Foundations	12 months
II	Chapter 2	Narrative Detection and Quantification across Macro, Micro and Market–Wide Sources	8 months
III	Chapter 3	Volatility Components Modeling and Benchmarking from High–Frequency Data	6 months
IV	Chapter 4	Narrative–Driven Volatility Structural Breaks, Causality and Forecasting	12 months
V	Conclusion	Narrative–Driven Risk Management Framework and Final Synthesis	3 months

- Stage I and part of Stage II and III have been completed during the Qualifier period.
- Thesis submission expected by July 2027, with defense preparation period reserved (Jul-Nov 2027).

Stage I – Data Collection & Literature Review (Introduction & Chapter 1)

Objectives:

- Establish conceptual and methodological foundations for narrative modeling in finance.
- Collect, clean, and structure textual and financial market data.
- Conduct a literature review to define financial narratives and identify methodological gaps.

Data Sources:

- RavenPack v1 news headlines via Quoniam partnership.
- Central-bank speeches (BIS Gigando, since 1996), corporate 10-K/10-Q filings (EDGAR) and Earning Call Transcripts (LSEG).
- Deutsche Börse High-Precision Timestamp databases (nanosecond trades and quotes since Feb. 2018) and LSEG (recent trades and quotes).
- Custom Python library for ingestion and Parquet/CSV storage, version control, and DMP compliance.

Outputs:

- Preprint hypothesizing about impact of alternative data on asset pricing.
- Systematic literature review.
- Private GitHub and Kaggle repositories.

Stage II – Key Financial Market Narratives (Chapter 2)

Goal: Detect, quantify, and compare macro, micro, and market-wide financial narratives using advanced NLP techniques.

Methodology:

- Three intensity score measurements: (1) supervised zero-shot latent embedding, (2) LLM-supervised, (3) unsupervised topic-modeling.
- Sentiment filtration: Mohammad, 2025 lexicon-based, Loughran and McDonald, 2011 lexicon-based, Araci, 2019 approach
- Multi-source text corpus: *news headlines* (market-wide), *central-bank speeches* (macro), *10-K/10-Q filings* and *earning calls transcripts* (micro).
- Evaluate explanatory power of narrative intensities on market variables (via rolling univariate regressions).

Outputs:

- Preprint on multidimensional aspect of polarity in text.
- Research paper employing supervised and unsupervised techniques for narrative modeling and quantifying.
- Private and Public GitHub and Kaggle repositories, including daily narrative scores.

Stage III – Market Microstructure & Volatility Modeling (Chapter 3)

Objectives:

- Build a high-frequency volatility modeling benchmark from Deutsche Börse nanosecond data.
- Build daily realized and 1-month implied volatility databases.
- Build rolling 3-months daily Hurst exponent and vol-of-vol databases.

Research Tasks:

- Model microstructure effects from UFTs and HFTs.
- Comparison of realized volatility estimators and jump detection tests.
- Estimation of implied vol surfaces, Hurst exponent and vol-of-vol.

Outputs:

- Research paper on stylized facts about High-Frequency Trading impact on financial markets.
- Preprint benchmarking various volatility estimators from High-Frequency Trading data.
- Private and Public Kaggle repositories, including daily volatility features.

Stage IV – Narrative-Driven Volatility Structural Breaks (Chapter 4)

Objectives: Integrate narrative information into the modeling and forecasting of volatility regimes.
Research Tasks:

- Structural Break Detection from Bai–Perron, CUSUM, and Kernel methods to volatilities, Hurst exponents, and vol-of-vol series.
- Rolling regressions and state-space models of narrative intensity on volatility dynamics.
- Causal Machine Learning frameworks to identify drivers of volatility regimes.
- Narrative-enhanced models for volatility and volatility roughness forecasting.

Outputs:

- Preprint assessing the explanatory power of narratives in volatility dynamics.
- Research paper on causal relationship between narratives and volatility, and narratives forecasting power.
- Private and Public Kaggle repositories, including daily volatility features.

Stage V – Conclusion

Objective: Synthesize the theoretical, empirical, and methodological insights gained across the four stages of the thesis.

Key Contributions:

- Establish a quantitative definition and measurement framework for financial narratives using advanced NLP methods across multiple textual domains.
- Link narrative to market volatility dynamics using high-frequency data.
- Demonstrate that narrative evolution anticipate structural breaks and volatility regime transitions.
- Provide new evidence of causal relationships between macro-, micro-, and market-wide narratives and volatility dynamics.

Outlook:

- Extend narrative–volatility models to cross-asset and global settings.
- Investigate the interaction between narrative diffusion, information flow, and volatility contagion.
- Develop narrative-based indicators for macro-financial monitoring and volatility investing.

Paper Overview

Title	Venue	Status	Date
Hypothesizing Multimodal Influence: Assessing the Impact of Textual and Non-Textual Data on Financial Instrument Pricing Using NLP and GenAI	**SSRN** [1]	Published	Feb 2025
An Algorithmic Framework for Systematic Literature Reviews: A Case Study for Financial Narratives	**Financial Innovation**	Under review	May 2025
TOPol: Capturing and Explaining Multidimensional Semantic Polarity Fields and Vectors	**arXiv** [2]	Published	Oct 2025
Nanosecond Microstructure: High-Frequency Traders Participation Stylized Facts	TBC	Ongoing [3]	Jan 2026
Quantifying Market-Relevant Narratives: A Comparative Framework for Narrative Modeling	TBC	Ongoing	Mar 2026
Benchmark of Realized Volatility Estimators from Nanosecond Data	**SSRN**		2026
Narratives and Volatility Dynamics in Equity Markets	**SSRN**		2027
Narratives as Causal Drivers of Volatility Regimes	TBD		2027

[1] DOI: 10.2139/ssrn.4698153
[2] DOI: 10.48550/arXiv.2510.25069
[3] Preprint submitted to the 17th Annual Hedge Fund Research Conference

PhD Thesis Cohesion

- **Unified Objective:** The thesis establishes an integrated framework linking financial narratives and volatility dynamics. It moves from the detection and quantification of narratives in textual data to their empirical connection with realized and implied volatility. Each chapter contributes to understanding how narratives emerge, evolve, and potentially drive structural changes in financial markets.
- **Progressive Methodological Depth:** The research follows a clear methodological progression. It begins with a systematic literature review and conceptual definition of financial narratives, advances to supervised and unsupervised NLP approaches for narrative detection, continues with high-frequency volatility modeling, and culminates in causal and predictive analyses of narrative–volatility interactions. This structure ensures increasing analytical depth and integration across disciplines.
- **Scientific and Practical Relevance:** The work contributes to both the theoretical foundations of narrative economics and its quantitative applications in finance. By combining NLP, econometrics, and machine learning, it enhances interpretability and predictive power in volatility modeling. The results have implications for alpha research, portfolio and risk management, policy making, and the broader understanding of information flow in financial markets.

Outline of Doctoral Education Programme

Approx. 15 EC's are planned to be earned before 2026 (remaining 15-20 EC's are scheduled to be earned before 2027):

- 20 scientific discipline EC: industry and academic conferences, PhD training schools, paper reviews, etc.
- 10-15 generic EC: PhD courses on academic activities, personal development, ML, etc.

Outline of Doctoral Education Programme

Scientific Discipline ECs

Title	EC	Location	Done
Vienna MSCA Conference	1.5	WU Vienna	✓
COST Action Coordination	3	BFH	✓
COST Action Meeting in Brussels	1	Brussels	✓
COST PhD Training School at UT	2.5	University of Twente	✓
Review paper for IEEE SDS2024	0.5	BFH	✓
QuantMinds 2025, 2026, 2027	4.5	London	
Conference Advances in Mathematical Finance	1.5	University of Freiburg	✓
Review paper for AAAI 2026	2	BFH	✓
Paper/Poster Presentation	2	Various locations	
23rd Winter school on Mathematical Finance	1	University of Amsterdam	
17th Annual Hedge Fund Research Conference	2	Université Paris Dauphine-PSL	
Obtained / Total	**12 / 21.5**	-	

Outline of Doctoral Education Programme

Generic ECs

Title	EC	Location	Done
Introduction to programming in C++	1.5	University of Twente	
PhD/EngD Introductory Workshop + Academic Integrity	1.5	University of Twente	
Academic Publishing	2	University of Twente	
Presentation skills	2	University of Twente	
Data Management	1	University of Twente	
Scientific Information	0.5	University of Twente	
Introduction to R	0.2	University of Twente	
Data visualization with R	0.2	University of Twente	
Narrative Economics by Robert Shiller	0.5	Coursera	
Statistical Learning with Python	1.5	Coursera	
Machine Learning Specialization by Andrew Ng	3.5	Coursera	
Obtained / Total	**0 / 14.4**	-	

Achievements

Project active participations:

- **COST Action 19130 on Fintech and Artificial Intelligence in Finance:** Co-coordinated events, internal communication, and budget.
- **MSCA Doctoral Network "Digital Finance":** Presented at event, admin tasks.

Industry collaborations:

- **Deutsche Börse:** Conducted a research internship focused on high-frequency trading classification and market microstructure analysis using nanosecond-level data.
- **Quoniam Asset Management:** Ongoing research internship on narrative-based alpha factor development using financial texts.

Achievements

Outputs:

- Various preprints and research papers.
- `realized-library` [4] [5]: Python package for realized volatility estimation and jump detection using nanosecond-level HFT data.
- `TOPol` [6]: transformer-based framework for Topic–Orientation Polarity shift detection to analyze semantic and polarity drift in texts.
- Literature discovery prototype: AI-enhanced tool for literature discovery and synthesis, integrating OpenAlex API, vector search, and graph-based semantic clustering and reranking.

[4] https://github.com/GabinTB/realized-library
[5] https://pypi.org/project/realized-library/
[6] https://github.com/GabinTB/TOPOL

Plan for the Remaining Period

Academic and Industry events:

- **PhD Training Schools:** active participation.
- **Conferences:** participation, paper submission, paper presentation.
- **Academic Collaborations:** visit, research papers.

Next short-term milestones:

- UTwente courses: C++ (ongoing), academic publishing (ongoing), presentation skills (ongoing), introductory workshop (Mar 2026), scientific information (Mar 2026), data management bootcamp (Jun 2026)
- Finalization and submission of HFT impact paper and market-wide narrative quantifying frameworks paper writing
- Completion of narrative detection framework (macro and micro levels).
- C++ improvement of `realized-library` package and volatility estimators benchmark.
- Online courses, conferences and training schools.

Overall focus: Consolidate interdisciplinary results, report the results in research papers, advertise research (conferences and social media).

Table of Contents

Proposed Thesis Outline (1)

Introduction: NLP for Financial Narrative Modeling

Chapter 1: Systematic Literature Review of Textual analysis and Narratives in Financial Markets

Chapter 2: Financial Narratives Detection and Processing

- Market-Wide Narratives (news headlines)
- Macro Narratives (central bank speeches)
- Micro Narratives (corporate filings and earnings calls)
- Comparative Analysis of Narrative–Market Relationships

Proposed Thesis Outline (2)

Chapter 3: Market Microstructure and Volatility Modeling

- Market Participants Behavior and Market Efficiency
- Realized and Implied Volatility Estimation
- Volatility Roughness and Vol-of-Vol Analysis

Chapter 4: Narrative-Driven Volatility Dynamics Forecasting

- Volatility Structural Break Detection
- Narrative–Volatility Dynamics
- Causality and Forecasting Frameworks

Conclusion: Synthesis of Narrative–Volatility Interactions and Future Research Directions

News Headlines

Source: RavenPack Analytics v1 dataset [7] (via Quoniam AM partnership).
Content: Global financial and macroeconomic news headlines enriched with metadata.
Structure:

- Core fields: id, publication time, source, and headline text.
- RavenPack metadata: sentiment scores, entity, relevance to entities, entity categories (companies, indices, commodities, currencies), and topic tags.
- Time coverage: 2000–present, event-level granularity.
- Used to build market-wide narrative tagging and daily narrative intensities.

[7] https://www.ravenpack.com/products/edge/data/news-analytics

10-K and 10-Q Filings

Source: SEC EDGAR database [8] (U.S. Securities and Exchange Commission).
Content: Company-specific financial disclosures (10-K annual reports and 10-Q quarterly reports).
Structure:

- Extracted sections: Item 1A – Risk Factors, Management Discussion and Analysis (MD&A), and Business Overview.
- Text parsed from structured files via Python API.
- Mapped to firm identifiers (ticker, CIK, ISIN) and time-aligned with market data.
- Used to extract micro-level narratives about firm-specific risks and strategic outlook.

[8] https://www.sec.gov/search-filings/edgar-application-programming-interfaces

Earnings Call Transcripts

Source: London Stock Exchange Group (LSEG) Transcripts database [9].
Content: Full-text earnings call transcripts, including management remarks and Q&A sessions.
Structure:

- Metadata: company ticker, ISIN, event date, speaker role (CEO, CFO, analyst).
- Extracted via LSEG Workspace API with unified formatting.
- Used to extract micro-level narratives about firm-specific risks and strategic outlook.

[9] https://www.lseg.com/en/data-analytics/financial-data/company-data/events/earnings-transcripts-briefs

Market Data and Asset Prices

Macroeconomic Data: St. Louis Federal Reserve (FRED API [10])

- Indicators: Fed Funds Rate, CPI, PPI, GDP, Unemployment, Nonfarm Payrolls.
- Monthly frequency, used for macroeconomic condition indices and regime labeling.

Live and Historical Market Data: LSEG Workspace API [11]

- Historical data covering many assets (trade prices, bid–ask quotes).
- Fundamental data (balance sheet, income statement, valuation ratios).
- Integration with corporate events, analyst estimates, and earnings call metadata.

High-Frequency Data: Deutsche Börse Marketplace [12] (Eurex & Xetra)

- Nanosecond-level timestamped trades and order-book messages (HPT & HPT-All).
- Instruments: equity and index spot, futures, and options.
- Enables participant classification (UFT, HFT, conventional) and market microstructure analysis.
- Used for realized and implied volatility estimation.

[10] https://fred.stlouisfed.org/docs/api/fred/
[11] https://developers.lseg.com/en/api-catalog
[12] https://console.marketplace.deutsche-boerse.com/home

Data Management

Data Collection and Storage:

- Data accessed via official APIs or institutional databases (RavenPack, LSEG, EDGAR, FRED, Deutsche Börse).
- Stored in structured (CSV, Parquet) and semi-structured (Markdown, JSON, PDF) formats.
- Secure storage on private servers with redundancy; eventually replication on private Kaggle repositories.

Data Preprocessing:

- Quality checks and alignment across textual and market datasets (timestamps, tickers, entities).
- Text preprocessing: cleaning, embedding generation (Transformers models).
- Market preprocessing: regular market hours filtering, trade–quote matching, microstructure event labeling.
- All pipelines version-controlled via private GitHub repositories with custom Python pipelines.

Narrative Definition

Proposition:

- Narratives are recurring, structured collections of financial topics that shape how market participants interpret and respond to market information.
- Each narrative is characterized by:
  - A set of semantically coherent topics (e.g., inflation, liquidity, policy uncertainty).
  - An intensity score capturing its relative prominence and emotional tone over time.
- The framework allows both cyclic narratives (e.g., inflation, monetary policy) and emerging narratives (e.g., AI, energy transition) to be modeled dynamically.

Financial Market Topics:

- Macroeconomics: Inflation, growth, GDP, unemployment, fiscal policy, recession risk.
- Geopolitics: Trade tensions, sanctions, wars, political uncertainty, global supply chains.
- Monetary Policies: Interest rates, quantitative easing, central bank communication, forward guidance.
- Assets: Equity markets, commodities, FX, credit, volatility, crypto assets.
- Etc.

Sentiment Analysis

Loughran and McDonald, 2011 Lexicon Approach:

- Finance-specific lexicon classifying words as positive, negative, uncertain, or litigious.
- Captures direction and uncertainty, commonly used in empirical asset pricing.

Mohammad, 2025 Lexicon Approach:

- Valence–Arousal–Dominance model: measures emotional valence (pleasantness), arousal (intensity), and dominance (control).
- Provides multidimensional sentiment representation.
- Applied on tokenized texts using weighted averages of VAD scores per text.

Araci, 2019 Embedding Approach:

- Transformer-based language model fine-tuned on financial text corpora.
- Outputs text-level probabilities for positive, neutral, and negative sentiment.

Narrative Modeling

Microeconomic narratives: 10-K, 10-Q and earning call transcripts (Flynn and Sastry, 2024);
Macroeconomic narratives: central bank speech transcripts from Feldkircher et al., 2021;
Overall narratives: Bybee et al., 2024 and "evergreen" extended narratives, reservoirs and super-narratives from Bhargava et al., 2023 and Lee et al., 2024:

- List of predefined financial narratives (e.g., "Inflation", "COVID-19", "US Growth", "Market Crash", etc.), classified in reservoirs and super-narratives.
- Represents a stable taxonomy of market stories, allowing consistent tracking but limited discovery of new ones.

Supervised Methods: Discover and track well-known cyclical and past temporary narratives:

- Similarity Tagging: Compute cosine similarity between text embeddings and narrative descriptions; assign top-matching narratives per text, aggregate at daily or intraday frequency to obtain narrative intensities.
- LLM Tagging: Use state-of-the art LLM with structured agentic framework to classify each text into predefined narratives; enables interpretability and narrative revalidation through self-consistent reasoning checks.

Narrative Modeling

Unsupervised method: Discover and track emerging narratives without predefined labels.

- Based on the LDA topic-modeling from Bybee et al., 2023;
- Graph-Based Topic Modeling: Enhancing the work of Grootendorst, 2022, the goal is to embed texts using sentence-transformers, reduce dimensionality with UMAP, and cluster via Leiden algorithm; each cluster represents a latent narrative evolving through time.
- Unsupervised LLM Approach: On daily batches of text, prompt an LLM to detect narratives and tag text.
- Advantages: Captures new market themes dynamically, providing early detection of evolving macro or micro narratives absent in predefined taxonomies.

Narrative Evolution

Intensity Scoring:

- Overall Intensity: Cumulative narrative-related text weights per day.
- Positive/Negative Intensity: Cumulative positive- or negative-sentiment narrative-related text weights per day.
- Super-Narrative Intensity: Cumulative narrative-related text weights per super-narrative per day.
- Reservoir Intensity: Cumulative narrative-related text weights per reservoir per day.

Temporal Tracking:

- Compute rolling correlations between narrative intensities and financial variables.
- Detect structural changes in narrative relevance using rolling univariate regressions.

Narrative Evolution - Super narratives

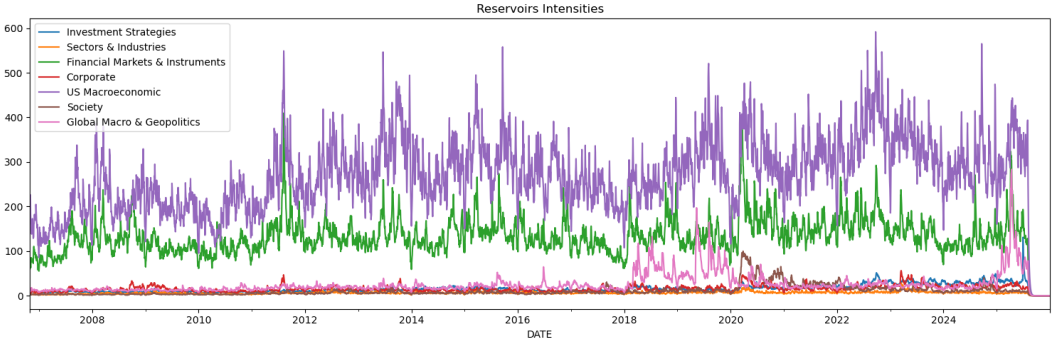

Figure: *Super Narratives Intensity Scores*

Narrative Evolution - Major temporary narratives

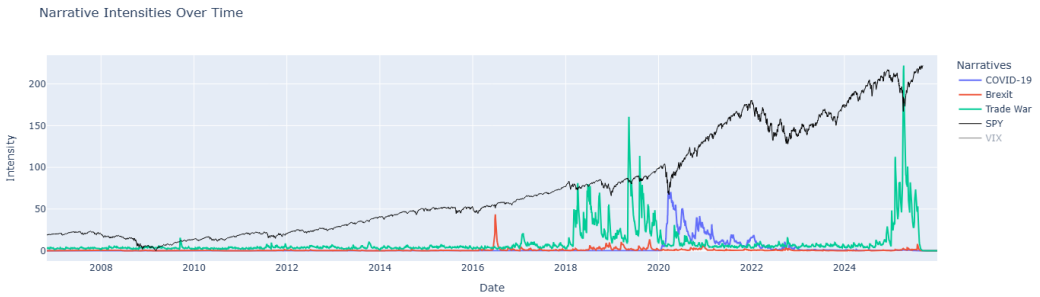

Figure: *Major Temporary Narratives Neg. Intensity Scores*

Narrative Evolution - Macroeconomic narratives

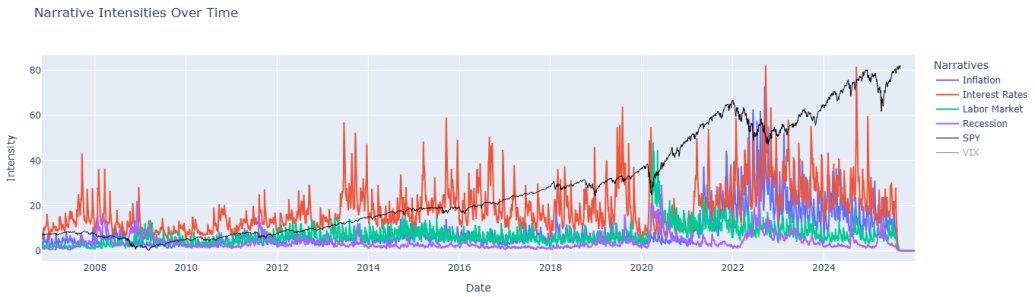

Figure: *Macroeconomic Narratives Neg. Intensity Scores*

Narrative Evolution - Market crash narratives

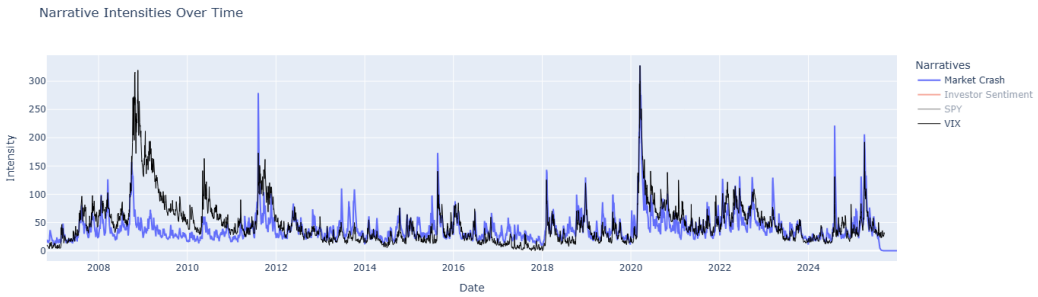

Figure: *Market Crash Narratives Neg. Intensity Scores*

Narrative Explanatory Power - S&P500

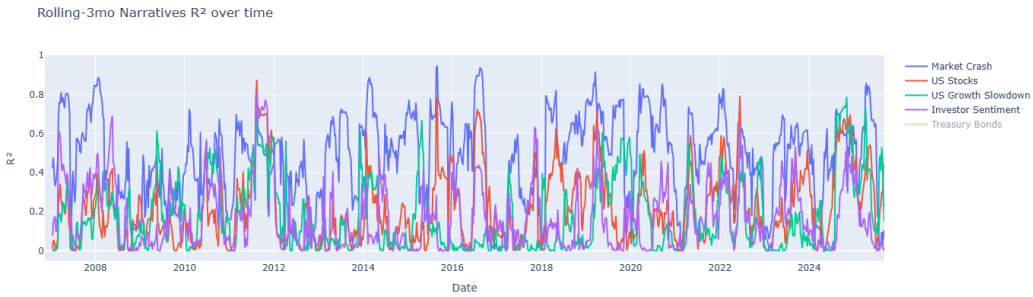

Figure: *Rolling r^2 of 3m-rolling regression between top narratives and $SPY*

Narrative Explanatory Power - S&P500

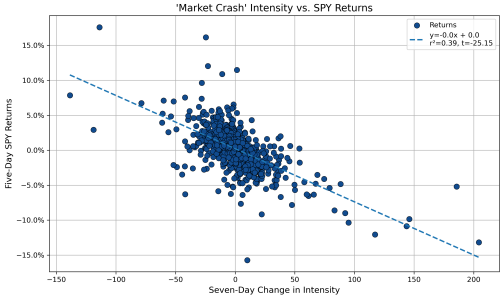

Figure: (a) Overall Regression Market Crash vs. $SPY

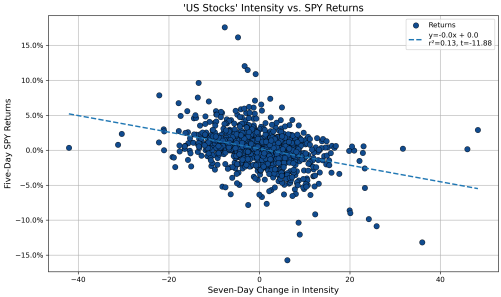

Figure: (b) Overall Regression US Stock vs. $SPY

Narrative Explanatory Power - VIX

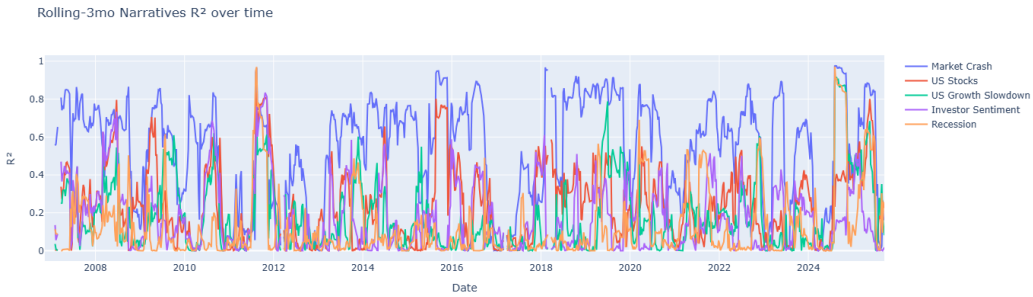

Figure: *Rolling r^2 of 3m-rolling regression between top narratives and $VIX*

Narrative Explanatory Power - VIX

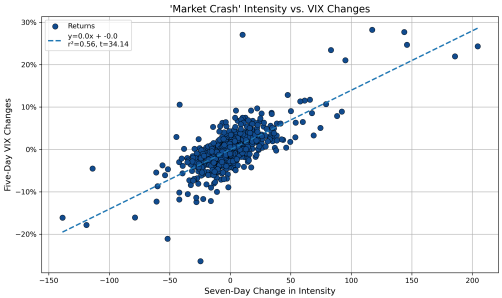

Figure: (a) Overall Regression Market Crash vs. $VIX

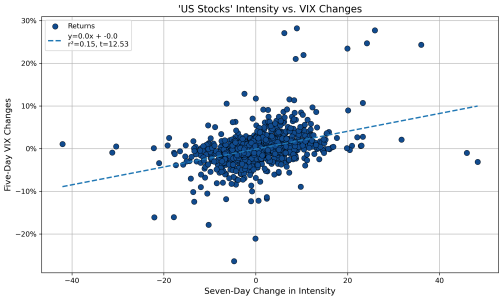

Figure: (b) Overall Regression US Stock vs. $VIX

Historical Realized Volatility

Objective: Compute jump- and noise-robust realized volatility measures from nanosecond-level HFT data.

- Developed the `realized-library`, a fast Python/C++ package for realized volatility estimation and price jumps detection.
- Use Deutsche Börse data (Xetra & Eurex, 2018–present) filtered to regular trading hours (09:30–17:00).
- Estimated daily volatility using multiple approaches: Realized Variance, MinRV, MedRV, Realized Kernel, etc.
- Construct daily time series of realized volatility, then derived realized Hurst exponents and volatility-of-volatility estimates.

Historical Implied Volatility

Objective: Extract option-implied volatility surfaces and higher-order volatility metrics.

- Use intraday snapshots of closest 1-, 3-, and 12-month maturities per underlying asset as reference options.
- Apply the Black–Scholes inversion to compute implied volatility from option prices (Eurex dataset).
- Select near-the-money contracts with sufficient liquidity and daily closing quotes for stability.
- Create daily close implied volatility timeseries and derive implied vol-of-vol.
- Use volatility surfaces to obtain time series of implied Hurst exponents.

Structural Breaks

Objective: Detect and interpret structural changes in volatility regimes through the lens of financial narratives.

- Apply multiple change–point detection techniques to volatility features: Bai–Perron multiple break test, CUSUM and Bayesian and Kernel-based methods for robustness.
- Identify volatility regime shifts (e.g., high– to low–vol regimes, smooth– to rough–vol transitions).
- Align detected breakpoints with changes in narrative intensities and sentiment indicators:
  - Overall (news headlines)
  - Macroeconomic narratives (policy, inflation, liquidity).
  - Micro narratives (corporate risk, earnings uncertainty).
- Evaluate whether narrative shocks precede or coincide with structural volatility transitions.

Output: Narrative–aligned map of volatility regime shifts, providing interpretability to purely statistical break detection.

Narrative Causality and Volatility Forecasting

Objective: Quantify and predict how narrative dynamics influence volatility structure and evolution.

- **Causality Analysis:**
  - Test causal direction between narrative intensities and volatility components (realized and implied, volatility, roughness, vol-of-vol).
  - Combine econometric and machine learning methods? E.g.: Granger causality, transfer entropy, Double Machine Learning, and Causal Forests.
- **Forecasting Framework:**
  - Integrate narrative features into volatility forecasting models.
  - Evaluate predictive gains versus baseline volatility models.
  - Explore narrative-timed volatility factor strategies (e.g., "short rough, long smooth" regimes).

Summary of Contributions

Thesis goal: Model financial narratives and volatility co-dynamics using NLP, econometrics, and high-frequency data.

- Establish a unified framework to quantify, track, and compare macro, micro, and market-wide financial narratives.
- Build high-frequency datasets and open-source tools (`realized-library`, `TOPol`) for volatility and semantic analysis.
- Develop a complete volatility modeling pipeline: realized, implied, roughness, and vol-of-vol.
- Connect narrative shifts to volatility and explain and forecast regime transitions.

Broader impact:

- Contributes to the empirical foundations of *Narrative Economics*.
- Enhances financial stability analysis, risk management, and macro-financial policy interpretation.

Future research directions:

- Extend the narrative–volatility framework to cross-asset and international markets.
- Model information diffusion and contagion through narrative propagation networks.
- Explore narrative-based volatility factors for portfolio construction and risk monitoring.
- Incorporate multimodal features (text, price, and sentiment) into deep forecasting architectures.

Conclusion

Thank you!

Questions and Discussion

Gabin Taibi
PhD Student, University of Twente
PhD Researcher, Bern University of Applied Sciences
gabin.taibi@utwente.nl, gabin.taibi@bfh.ch