An Algorithmic Framework for Systematic Literature Reviews

A Case Study for Financial Narratives

A Quantitative Finance Perspective

Gabin Taibi Joerg Osterrieder

BFH Bern University of Applied Sciences

March 27, 2026

Notes — Slide 1

Welcome. This talk presents a dual contribution: a reproducible algorithmic framework for systematic literature reviews and its application to the rapidly growing field of financial narratives. 20 minutes, happy to take questions at the end.

The Narrative Turn in Finance

- **Shiller (2017, 2019):** Economic narratives go viral, drive booms, panics, recessions — "narrative economics" as a research program
- **Grossman & Stiglitz (1980):** Information heterogeneity persists in equilibrium — narratives fill the gap between prices and beliefs
- **Bybee et al. (2023):** News-based narrative factors explain cross-sectional risk premia beyond traditional factor models
- **Bybee et al. (2024):** News attention tracks business cycles, forecasts aggregate stock returns out-of-sample
- **Flynn & Sastry (2024):** Belief-driven macroeconomic fluctuations extracted from firm disclosures

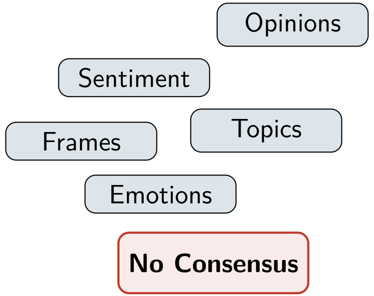
Narratives have measurable, quantifiable effects on financial markets.

Notes — Slide 2

Narratives have measurable, quantifiable effects on markets — not soft science. Shiller's Nobel lecture made the case. Since then, quantitative evidence has accumulated rapidly. Bybee and coauthors show narrative factors rival Fama-French factors in explanatory power.

The Problem — Fragmentation

- "Financial narrative" means different things to different researchers
- Most studies reduce narratives to **sentiment** — a lossy compression
- Methods range from dictionary-based to LLM-based — **no consensus**
- No unified typology, no standard benchmarks
- Hard to build on each other's work, hard to compare results



Notes — Slide 3

The field is growing fast but incoherently. This motivates a systematic review. Everyone uses the term "financial narrative" but they mean very different things. Most default to sentiment, which discards structure, causality, and temporal dynamics.

Research Questions & Contribution

Research Questions

RQ1 How can NLP be used to **quantify and model** financial narratives?

RQ2 How is “financial narrative” **defined and applied** in the literature?

Dual Contribution

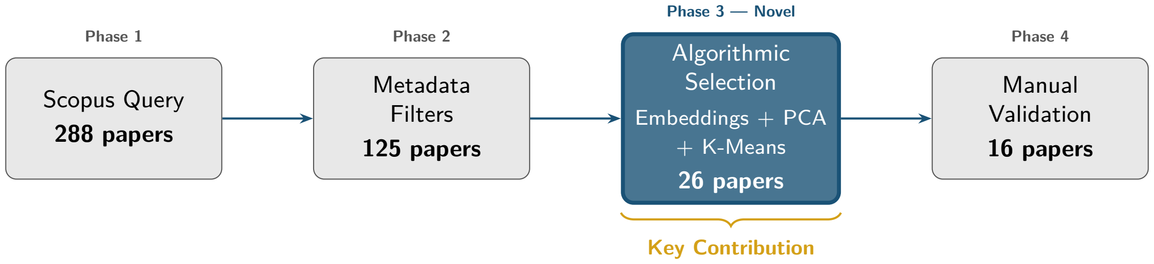
(a) **Algorithmic SLR Framework** — reproducible, embedding-based paper selection replacing subjective screening

(b) **Structured Synthesis** — concept matrix revealing gaps, two-stream taxonomy, future directions

Notes — Slide 4

The methodology IS a contribution, not just scaffolding. Most SLRs treat their method as boilerplate. Ours is reusable, quantitative, and addresses the scalability problem of systematic reviews in fast-moving fields.

SLR Pipeline Overview



- **Phase 3** replaces subjective screening with semantic embedding similarity
- 288 → 125 → 26 → 16: systematic, reproducible reduction

Notes — Slide 5

Phase 3 is the novel contribution. Everything else is standard SLR methodology per PRISMA. The key innovation is replacing human judgment in screening with embedding-based semantic similarity and unsupervised clustering.

The Algorithmic Core (Phase 3)

Semantic Scoring

1. 6 screening statements formulated from RQs
2. Paraphrased via GPT-3.5 for robustness
3. Embedded via `text-embedding-3-small`
4. Each abstract embedded identically
5. **Cosine similarity** $\rightarrow$ 6 scores per paper

Dimensionality Reduction & Clustering

1. Z-score normalization across scores
2. **PCA** (KMO $= 0.815$, Bartlett $p < 0.001$)
3. **K-Means** clustering ($k = 3$)
4. High-relevance cluster identified
5. $\rightarrow$ **26 papers** selected

Abstracts → Embed → Cosine Sim → Z-Score → PCA → K-Means → **26 Papers**

Notes — Slide 6

Semantic matching, not keyword matching. The embedding captures meaning. We use cosine similarity between paper abstracts and screening criteria in embedding space. This is more nuanced than Boolean keyword search — it captures paraphrases, related concepts, and semantic proximity.

Stability & Robustness

PCA Variance Thresholds

- Tested 80%, 85%, 90%, 95%, 98% explained variance
- **All thresholds yield the same 26 papers**
- Result insensitive to PCA hyperparameter

Conclusion

Results robust to methodological choices

Framework produces stable outputs across reasonable parameter ranges

Clustering Algorithm Comparison

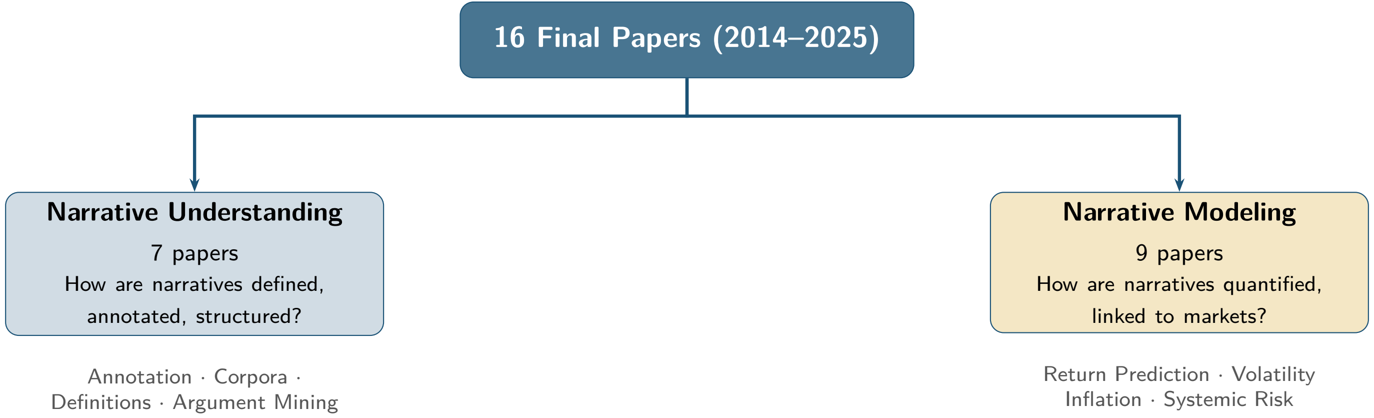
Method	**Silhouette**	**Cluster Size**
K-Means	Best	Balanced
GMM	Lower	Unbalanced
Agglomerative	Lower	Unbalanced

- K-Means: best balance of cluster size, relevance, and Silhouette score
- GMM and Agglomerative produce skewed clusters

Notes — Slide 7

Addresses the obvious quant concern: is this sensitive to hyperparameters? No. We tested multiple variance thresholds and three clustering algorithms. K-Means consistently produces the most interpretable and balanced result. The 26-paper set is stable.

Two Streams of Literature



Notes — Slide 8

Understanding = “what is a narrative?” Modeling = “what can narratives predict?” This two-stream taxonomy emerged from the concept matrix. Understanding papers build the linguistic and conceptual infrastructure. Modeling papers build the predictive applications.

Narrative Understanding — Highlights

- **Hu et al. (2021):** Annotation scheme for opinion and emotion in Chinese economic texts — one of the first structured frameworks
- **Zmandar et al. (2022):** *CoFiF Plus* — large-scale French financial narrative corpus with multi-task annotations
- **Sy et al. (2023):** BERT ensembles for **argument mining** in earnings calls — moves beyond sentiment to argumentative structure
- **Liu et al. (2024):** *Financial-STS* — detecting semantic drift in corporate disclosures via sentence similarity
- **Roos & Reccius (2024):** Rigorous conceptual definition —

Narratives $\neq$ topics $\neq$ sentiment

Narratives are structured, evaluative, temporally ordered expressions

Notes — Slide 9

Narratives are structured, evaluative expressions — not just positive or negative. Roos and Reccius provide the most rigorous definition: a narrative has temporal ordering, causal claims, and evaluative stance. This distinguishes it from topics, which are just co-occurring words, and sentiment, which is a scalar projection.

Narrative Modeling — Highlights

- **Tuckett et al. (2014):** Conviction narratives, regime detection in FX markets via directed sentiment
- **Chen et al. (2022):** COVID narrative virality $\rightarrow$ realized volatility transmission across markets
- **Borup et al. (2023):** Narrative expectations index **outperforms** sentiment indices for return prediction
- **Ma et al. (2024):** Narrative Energy Index (NEI) predicts sector and aggregate market returns
- **Hong et al. (2025):** 880K WSJ articles $\rightarrow$ narrative factors **outperform macro models** for inflation forecasting
- **Miori & Petrov (2023):** GPT + graph theory $\rightarrow$ narrative fragmentation as a proxy for **systemic risk**

Each study represents a potential **alpha signal** or **risk indicator**:
returns, volatility, inflation, systemic risk

Notes — Slide 10

Each is a potential alpha signal or risk indicator. Returns, volatility, inflation, systemic risk. The key takeaway: narratives are not just descriptive color — they have predictive power across asset classes and economic variables. Hong et al. is particularly striking: text beats macro data for inflation forecasting.

Concept Matrix

Paper	Theory			Context				Method			
	Sentiment	Structured	Formal Def.	Equity	Macro	Crisis	Other	Lexicon	Topic Model	Transformer	LLM
Tuckett et al. (2014)	■						■	■			
Hu et al. (2021)	■	■			■						
Chen et al. (2022)	■			■		■			■		
Zmandar et al. (2022)		■					■			■	
Borup et al. (2023)	■			■	■				■		
Sy et al. (2023)		■		■						■	
Miori & Petrov (2023)		■				■					■
Liu et al. (2024)		■		■						■	
Ma et al. (2024)	■			■	■				■		
Roos & Reccius (2024)		■	■				■				
Hong et al. (2025)					■				■	■	
Coverage	*5/11*	*6/11*	*1/11*	*5/11*	*4/11*	*2/11*	*3/11*	*1/11*	*4/11*	*4/11*	*1/11*

■ = covered   = **gap / opportunity**

Notes — Slide 11

The power of this table is in what is MISSING. The white space is the opportunity. Only 1 paper provides a formal definition. Only 1 uses LLMs. Only 2 study crisis contexts. The intersection of "formal definition + transformer + equity" is completely empty. These gaps directly motivate the research agenda on Slide 12.

Four Research Gaps

G1 No integration of structured representations + semantic methods + financial modeling

Papers either define narratives rigorously OR model them quantitatively — never both

G2 Contextual embeddings largely absent from financial narrative applications

Most modeling papers use topic models or lexicons; transformers appear only in understanding papers

G3 Empirical scope is narrow

Few asset classes beyond equities, limited data sources beyond news and filings

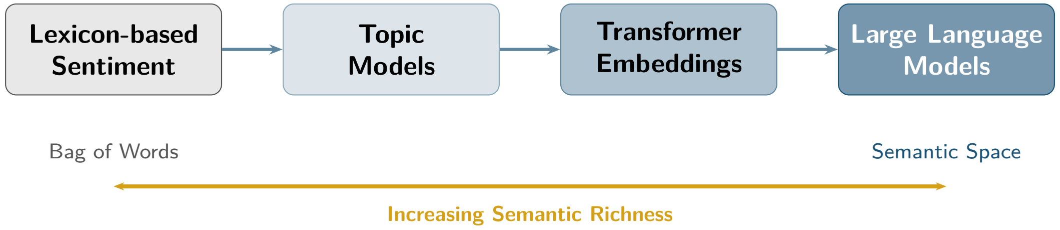
G4 Narratives treated as exogenous, never as endogenous market components

No paper models how narratives form, spread, compete, and reshape markets

Notes — Slide 12

Frame each gap as an opportunity for the audience. G1 is the biggest: nobody has combined all three elements. G4 is the most ambitious: treating narratives as endogenous requires agent-based or general-equilibrium modeling. Each gap is a viable research program.

From Sentiment to Semantic Space



Key Insight

Narratives ≈ **trajectories in semantic embedding space**

Position · Velocity · Clustering · Divergence

Familiar mathematical framework, fundamentally new input signal

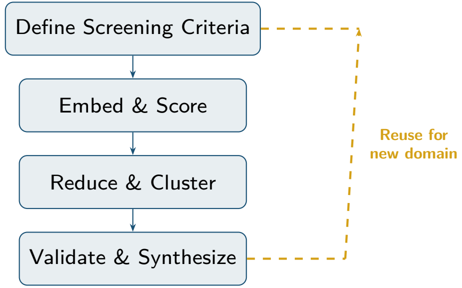
Notes — Slide 13

Trajectories in embedding space are analogous to trajectories in factor space. Familiar math, new input. A narrative is not a number — it is a path through a high-dimensional space. Its position tells you what is being discussed. Its velocity tells you how fast attention is shifting. Its clustering tells you consensus vs. fragmentation. Quants already know how to work with these objects.

A Reusable SLR Framework

Design Principles

- Works for **any fast-moving research domain**
- Semantic relevance assessment, not lexical matching
- Quantitative quality metrics replace subjective judgment
- **Reproducible**: same inputs → same outputs
- **Scalable**: handles hundreds to thousands of papers



Applicability

Climate finance · RegTech · Crypto assets · ESG · Any domain with rapid publication growth

Notes — Slide 14

Brief — secondary contribution but important for anyone doing literature surveys. The framework is domain-agnostic. Swap out the screening criteria and it works for climate finance, RegTech, crypto, or any other fast-growing field. The key advantage over manual screening is reproducibility and scalability.

Limitations

- **Single database:** Scopus only — may miss preprints (arXiv, SSRN) and grey literature
- **Abstract-level screening:** Full-text analysis might recover papers whose abstracts understate relevance
- **Closed-source embedding model:** OpenAI `text-embedding-3-small` — reproducibility depends on API stability
- **Medium-relevance cluster excluded:** Conservative choice; some relevant papers may sit at the cluster boundary
- **Temporal snapshot:** Literature evolving rapidly — results reflect the state at time of query

All limitations are **addressable** in future iterations — multi-database, full-text, open embeddings

Notes — Slide 15

Quants respect transparency about limitations. Every one of these is fixable. Multi-database queries are straightforward. Full-text analysis requires more compute but no methodological change. Open-source embeddings like those from Hugging Face can replace the OpenAI model. We chose pragmatism over perfection for the first iteration.

Future Directions & Call to Action

1. **Unified narrative typologies**
   Agree on what a "financial narrative" is — operationalize Roos & Reccius (2024)
2. **Cross-market, cross-asset studies**
   Beyond equities + news: commodities, fixed income, derivatives, social media, earnings calls
3. **Joint narrative–market models**
   Link narrative measures to quantitative market data in unified econometric or ML frameworks
4. **Open benchmarks for narrative extraction**
   Shared datasets, annotation standards, leaderboards for financial NLP
5. **Narratives as endogenous**
   Model formation, evolution, competition, and market impact jointly — the frontier

Notes — Slide 16

These are specific, actionable research programs. Each one could be a PhD thesis or a multi-year project. Item 5 is the most ambitious and the most important: treating narratives not as exogenous inputs but as endogenous components of the market system. This requires new modeling paradigms — agent-based, network-based, or general-equilibrium approaches.

Summary

What We Built

- **Framework:** Reproducible, scalable, semantic SLR pipeline
- **Application:** Structured synthesis of financial narrative literature

What We Found

- Field advancing rapidly but **fragmented**
- Two streams: **understanding** (what?) vs. **modeling** (what predicts?)
- Four clear research gaps with actionable directions

Key Message

Narratives are **central** to financial markets, not peripheral.

Thank You!

Gabin Taibi
gabin.taibi@bfh.ch

Joerg Osterrieder
joerg.osterrieder@bfh.ch

Paper available upon request

Notes — Slide 17

Crisp summary. Invite questions. The paper is available and we welcome collaboration on any of the future directions discussed. Thank you for your attention.

Backup: Key References

Foundational

- Shiller, R. (2017). Narrative Economics. *AER*.
- Shiller, R. (2019). *Narrative Economics*. Princeton UP.
- Grossman, S. & Stiglitz, J. (1980). On the Impossibility of Informationally Efficient Markets. *AER*.

Narrative Factors

- Bybee, L. et al. (2023). Business News and Business Cycles. *JF*.
- Bybee, L. et al. (2024). Narrative Asset Pricing. *JFE*.
- Borup, D. et al. (2023). The Anatomy of Sentiment-Driven Fluctuations. *JFQA*.

Modeling & Applications

- Hong, T. et al. (2025). News and Inflation Expectations. *WP*.
- Miori, L. & Petrov, S. (2023). GPT-Based Narrative Risk. *WP*.
- Ma, Y. et al. (2024). Narrative Energy Index. *IRFA*.
- Flynn, J. & Sastry, K. (2024). Belief-Driven Fluctuations. *WP*.

Definitions & Methodology

- Roos, M. & Reccius, M. (2024). Narratives in Economics. *JEL*.
- Paul, J. & Criado, A. (2020). Art of Writing Literature Reviews. *APJM*.

Notes — Slide 18

Backup references for Q&A. These are the most-cited papers in the review and the methodological anchors for the algorithmic framework.