QUANT RESEARCH

Narrative Contagion in Markets: How Attention Regimes Shape Semiconductor Volatility

This report explores how collective attention and market narratives help explain bursts of volatility in semiconductor equities, above and beyond fundamentals, by combining narrative-based indicators with a machine learning volatility model. As Robert J. Shiller wrote, *"Narratives 'go viral' and spread far, even worldwide, with economic impact."*

Martin Carrasco
martintuan.carrasco@alumni.esade.edu
Team Lead & Research Analyst, ESFS
Francesco Palma
francesco.palma@alumni.esade.edu
Research Analyst, ESFS
Angel Brendel
angelpaulnicolasmarcemile.brendel@alumni.esade.edu
Research Analyst, ESFS

CONTENTS PAGE

Executive Summary

Financial markets are not driven solely by fundamentals such as cash flows, interest rates, or macroeconomic conditions, but are also shaped by dominant narratives that influence how information is interpreted and which signals attract investor attention. Narratives organise complex economic realities into simplified stories, guiding expectations and coordinating behaviour across market participants. As a result, periods of narrative dominance or rapid narrative change may generate nonlinear market responses that are not easily captured by traditional asset pricing or forecasting models.

A narrative is not just "sentiment", it is a storyline that selects which facts matter, links them into a causal account, and makes certain interpretations easier to adopt quickly. The narrative-economics view is that such stories can spread and coordinate behaviour (Shiller, 2017). A complementary attention-based view formalises why this coordination can arise: attention is scarce, so investors often process sector-level signals more than firm-level detail, which can strengthen comovement and amplify sector moves (Peng & Xiong, 2006).

The semiconductor sector provides a particularly suitable environment to study narrative-driven market dynamics. Semiconductors lie at the core of modern economic growth, underpinning several developments (AI, cloud computing…), while simultaneously being exposed to geopolitical risks, supply chain disruptions, and policy interventions. These characteristics make the sector highly sensitive to shifts in attention and sentiment, with narratives related to technological disruptions, export controls, and many more.

This debate raises a central question :

In this paper, we use sector-level market data (SOXX), macroeconomic controls, and a narrative dataset to capture investor attention across multiple thematic dimensions and to construct measures of narrative attention, narrative concentration, and narrative shocks. Then, we integrate these measures into a machine learning model (XGBoost) to forecast next-month realised volatility at a monthly horizon. Importantly, we track not only "how much attention" there is, but also whether that attention is concentrated (low entropy) or fragmented (high entropy). Importantly, machine learning is employed not as a black-box forecasting tool, but as an explanatory device that allows for nonlinear interactions and state-dependent effects.

Esade Student Finance Society

From Sentiments to Narratives

Conceptual framework

Before defining the concept formally, consider a simple example. If many investors believe that “AI will drive the next wave of growth in technology,” this story will shape how they interpret news. Positive AI developments will receive strong attention, while unrelated signals may be underestimated. This kind of shared interpretive story is what we call a “narrative”.

Narratives shape what investors pay attention to, how they understand uncertainty, and how they form expectations about the future. Instead of treating narratives as direct predictors of prices or volatility, this study views them as shaping the context in which markets react. For instance, under a strong AI growth narrative, investors may react very positively to innovation news, while the same news might have a weaker impact when that narrative is absent. This section develops this idea by introducing narratives as “attention filters”, by measuring how concentrated narratives are using entropy, and by outlining testable implications for market dynamics.

Narratives as attention filters

Financial markets generate an enormous amount of information, including macroeconomic releases, firm-level news, policy decisions, and geopolitical events. However, investors have limited cognitive capacity and cannot process everything at once. As a result, attention becomes a scarce resource, and narratives play a key role in directing where that attention goes.

When a narrative becomes dominant, it acts like an attention filter. It highlights certain signals while pushing others into the background. For example, a strong narrative around artificial intelligence driving semiconductor growth may lead investors to focus heavily on innovation-related news while reacting less to macroeconomic tightening.

This filtering role has 2 major effects:

First, it increases coordination among investors. When many market participants focus on the same story, their expectations and actions become more aligned, leading to stronger and more persistent market movements.

Second, dominant narratives can make markets more fragile. By crowding out alternative interpretations, they reduce the diversity of beliefs, making markets more vulnerable to sharp reversals if the narrative is challenged or breaks down.

These dynamics are particularly visible in the semiconductor sector. The industry is often shaped by powerful narratives around technological leadership, supply chain resilience, and geopolitical rivalry. Such narratives can rapidly shift attention across firms, technologies, or regions, amplifying market reactions even when underlying fundamental changes are relatively modest.

Quantifying Narratives

Narrative concentration and Entropy

To move from an intuitive idea of narrative dominance to something measurable, we need a way to quantify how narratives are distributed over time.

This study captures this by modelling the narrative environment as a distribution of attention across multiple sector-related narratives. In simpler terms, it looks at how investor attention is divided across different stories at any given point in time.

These narratives may refer to firm-specific developments, technological trends, industry-wide themes, or geopolitical and policy issues.

Narrative concentration is measured using entropy, a concept from information theory that captures how spread a distribution is. When most attention is focused on a small number of narratives, entropy is low, indicating narrative dominance. When attention is spread more evenly across multiple narratives, entropy is high, indicating narrative fragmentation.

Entropy definition:

$$H = -\sum_{i=1}^{n} p_i \log p_i$$

During low entropy periods most investors focus on the same core story driving the sector. This shared focus leads to more aligned expectations and behaviour, which can reinforce market trends. However, it can also make markets more fragile, as a widely shared narrative increases the risk of sharp adjustments if it weakens or is challenged.

In high entropy periods, attention is spread across multiple competing stories, suggesting less agreement about the sector's outlook. In such settings, market reactions may be less coordinated but more unstable, as expectations shift more frequently.

Importantly, entropy is not a measure of optimism or pessimism. Instead, it captures how attention is structured. A rise in entropy does not necessarily imply negative sentiment, but rather a loss of narrative consensus. This distinction helps explain why entropy can influence volatility even when overall sentiment appears stable.

By tracking both the level and the change in narrative entropy over time, the framework captures not only stable narrative environments but also transitions between them. Rapid shifts in entropy indicate moments when the narrative landscape is being reconfigured, which may contribute to periods of market instability.

Quantifying Narratives

Testable implications

This framework leads to several testable ideas about how narratives, attention, and market behaviour are connected. These ideas are state-dependent, meaning that narratives do not influence markets all the time in the same way. Instead, their impact becomes stronger under certain conditions.

First, when investor attention is very high and focused on only a few dominant narratives, market reactions are expected to become stronger. When many investors are paying attention to the same stories, price movements can become larger and more volatile.

Second, sudden changes in narrative entropy are expected to occur before spikes in volatility or market declines. A sharp increase in entropy means that the shared narrative is breaking down and investors are becoming less aligned. A sharp decrease means that investors are quickly converging around a new dominant story. Both types of shifts reflect unstable expectations and may lead to higher uncertainty in markets. For this reason, changes in entropy may be more informative than entropy levels alone.

Third, narratives are expected to act as amplifiers rather than constant predictors of market outcomes. Unlike traditional risk factors, narratives are not assumed to have a steady, predictable effect on volatility or returns. Instead, their influence depends on how attention is distributed at a given moment and how narratives interact with broader market conditions. This means narrative measures may improve forecasting mainly by helping detect periods where markets are more likely to react strongly, even if they do not always improve average predictive accuracy.

These ideas guide the empirical strategy of the study. By including measures of attention and entropy in a machine learning framework, the analysis can capture complex relationships that standard linear models may miss. Using explainable machine learning also helps identify the conditions under which narratives matter most, linking the empirical results back to the conceptual framework.

In summary, this framework treats narratives as structural features of the information environment that shape how attention is distributed and how markets react. By focusing on how narratives concentrate, fragment, and shift over time, the approach moves beyond simple sentiment measures and provides a structured way to study how narratives contribute to volatility dynamics in the semiconductor sector.

Data Construction & Empirical Setting

To study whether narratives influence markets, we built a dataset combining three layers of information: market behaviour, macroeconomic conditions, and investor attention.

Market dynamics are represented using the SOXX semiconductor ETF, from which we compute monthly returns and realised volatility. Volatility is calculated using daily price fluctuations within each month and then annualised, so it reflects a standard financial risk measure rather than raw price changes.

We then include macro-financial controls such as the policy rate, the two-year yield, inflation, industrial production, unemployment, and credit spreads. These variables capture the economic environment and ensure that any detected narrative effect is not simply macroeconomic noise.

Finally, we construct narrative indicators using online attention data linked to semiconductor-related themes. Instead of measuring whether news is positive or negative, we measure how strongly investors focus on particular topics, allowing us to quantify shifts in collective attention. To move from a qualitative notion of narrative dominance to a quantitative framework, it is necessary to formalise how narratives are distributed over time. This study does so by modelling the narrative environment as a distribution of attention across multiple sector related narratives. These narratives may correspond to firm specific themes, technological developments, industry wide trends, or geopolitical and policy related issues.

Narrative concentration is measured, as seen previously, using entropy, a concept originating from information theory that captures the dispersion of a probability distribution. In this context, entropy measures how evenly attention is distributed across narratives at a given point in time. When attention is highly concentrated on a small number of narratives, entropy is low, indicating narrative dominance. When attention is more evenly spread across narratives, entropy is high, indicating narrative fragmentation (see definition in part “Quantifying Narratives”, page 5-6).

Because all sources must overlap, the final dataset covers August 2015 to December 2025 and contains 81 monthly observations. While relatively small, this structure is well suited for tree-based models designed to extract nonlinear relationships from dense feature sets (see GitHub link in Appendix)

Modelling Approach & Empirical Comparison

Explainable Machine Learning as an Economic Tool

We use a Random Forest model to capture nonlinear interactions between narratives and financial variables. A Random Forest builds many decision trees and averages them, allowing the model to identify conditional relationships automatically rather than imposing a fixed formula.

This is useful because narratives are unlikely to have a constant effect. Their impact depends on context. For example, whether investor attention is already elevated or fragmented.

Importantly, the model is not used to generate trading signals. Its purpose is explanatory: to test whether narrative structure contains incremental information about future volatility.

We train the model using data prior to 2021 and evaluate performance out-of-sample afterward to simulate a real forecasting environment.

Empirical comparison: Do narratives add information?

To evaluate the role of narratives, we compare two models:

1) A baseline model using only market and macroeconomic variables

2) A narrative-augmented model that adds attention, entropy, and narrative shock measures

Performance is assessed using Mean Absolute Error (MAE) and out-of-sample R^2.

Model	Out-of-Sample R^2	Mean Absolute Error (MAE)
Baseline (Market + Macro Only)	−0.178	0.0831
Narrative-Augmented Model	−0.189	0.0817

At first glance, the R^2 does not improve, which is common when forecasting volatility at monthly horizons. However, the MAE declines, meaning the model with narratives makes smaller prediction mistakes in absolute terms.

This is economically meaningful because risk management is concerned with reducing large forecast errors rather than maximising statistical fit.

Results: Narratives as Episodic Amplifiers

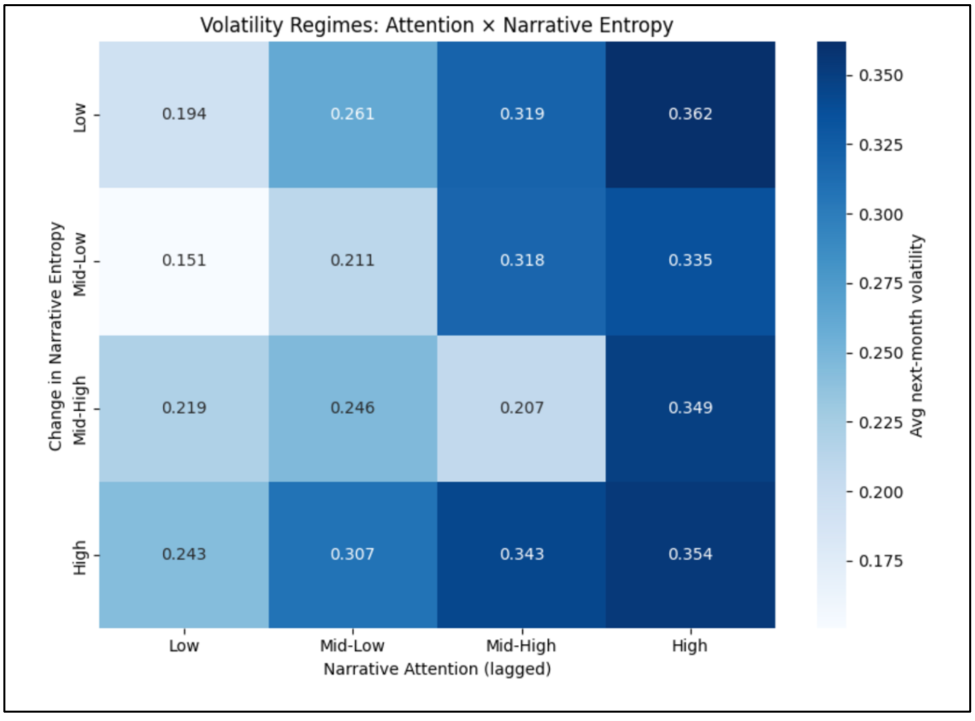
The empirical results show that narratives do not behave like continuous risk factors. Instead, they act as episodic amplifiers that matter only in specific regimes.

Narrative shock periods, defined as months with unusually high attention and strong narrative concentration, occur only 4 times out of 81 observations, confirming that narrative contagion is rare by construction.

Yet during these episodes, average annualised volatility rises from approximately 27.1% in normal periods to 30.5% in narrative-shock regimes, indicating a meaningful increase in market instability.

This supports the hypothesis that narratives amplify existing uncertainty rather than continuously driving it.

Figure 1: Volatility Regimes: Attention × Narrative Entropy" heatmap



The figure shows that volatility is highest when attention is high and entropy is low (when investors are collectively focused on one dominant story rather than multiple competing interpretation).

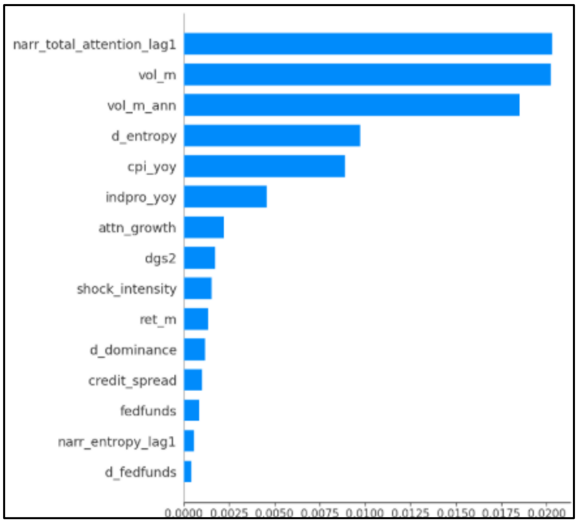
This pattern reinforces the interpretation of narratives as coordination mechanisms that temporarily synchronise expectations and increase fragility.

SHAP Analysis

Interpreting the Mechanism with SHAP Analysis

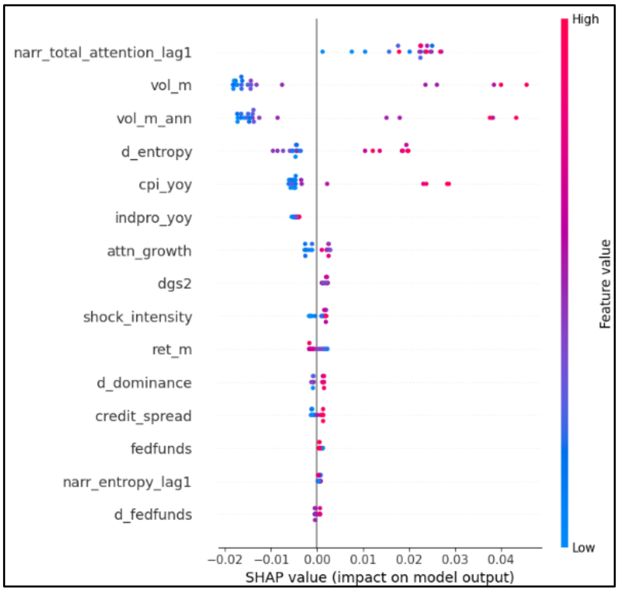
SHAP is a framework that applies cooperative game theory (Shapley values) to machine learning, treating features as players in a game whose payoff is the model prediction, and assigning each feature a fair contribution score. In plain terms, SHAP means: “For this month’s volatility forecast, what did the model treat as the main drivers?”

Figure 2: SHAP Feature Importance



Model interpretability is assessed using SHAP values. Lagged volatility remains the dominant predictor, validating the model’s internal consistency. Importantly, narrative attention and entropy changes rank among the most influential non-market drivers.

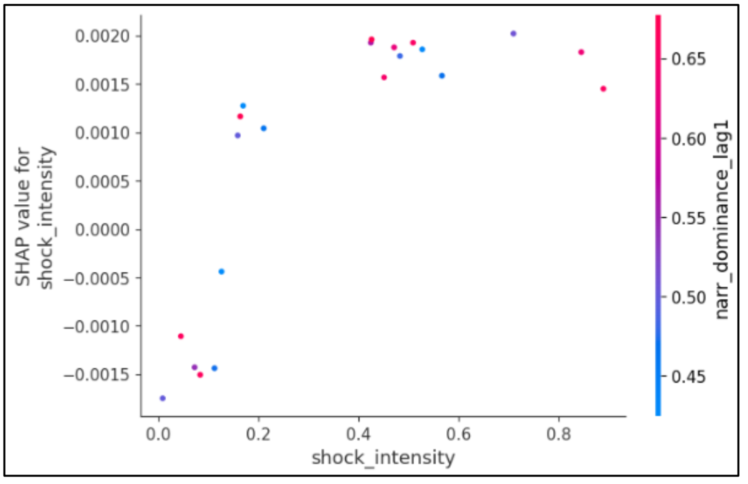
Figure 3: SHAP Summary



The main interpretive point is interaction: narrative variables matter most when they line up with a synchronised focus. Our SHAP summary states that high attention combined with low entropy (attention concentrated in a small set of themes) increases predicted volatility far more than similar attention levels when narratives are fragmented.

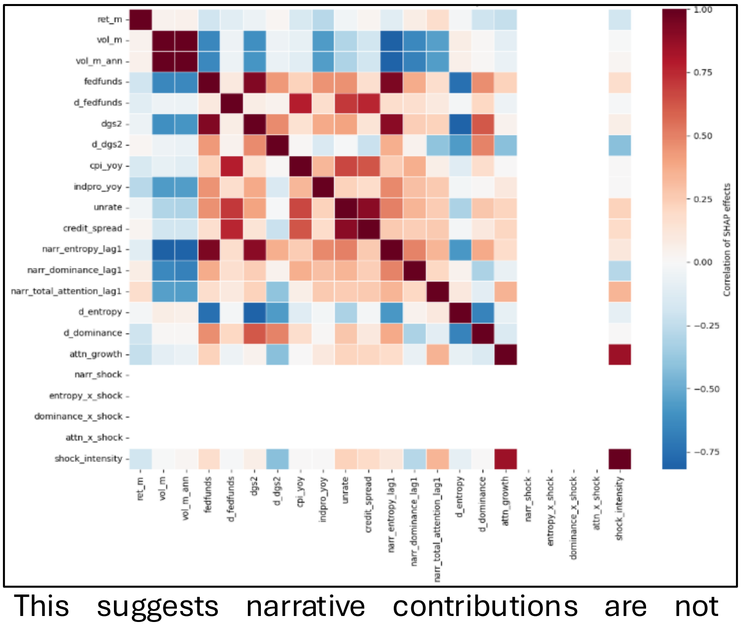
SHAP Analysis

Figure 4: SHAP Interaction Plot



Overall, the plot is consistent with an amplification story: higher shock intensity is associated with a more positive contribution to the model's predicted volatility, especially when narrative dominance is high.

Figure 5: SHAP Effect Correlation Heatmap



The heatmap shows correlations between features' SHAP effects, with blue indicating negative correlation and red positive, and a strong positive diagonal by construction. Warm blocks among macro controls suggest their SHAP contributions often move together, implying partial redundancy in effects (several inputs reflect the same broad state). Narrative-related variables sit in a looser cluster with mixed signs against the macro block and cool patches against volatility-lag features. This suggests narrative contributions are not simply duplicates of "volatility persistence". Shock_intensity's SHAP effects appear most positively aligned with attention/shock-related terms and less tightly linked to many macro controls, placing narrative/shock features as a partly distinct channel alongside standard controls.

Implications for Quantitative Analysis

Our findings shows that narratives behave more like state variables for diagnosing risk and regimes than like standalone return signals. This is consistent with our modelling objective: interpretability serves as explanation and risk understanding, not trading optimisation, and reports only modest average forecast improvement.

It is also consistent with how attention is measured in related work: online information demand can proxy “what investors are looking at” in a way that complements price and volume measures (Da, Engelberg & Gao, 2011). Finally, it matches how news-based uncertainty measures are used in macro-finance. They are typically treated as diagnostic indicators of uncertainty regimes rather than simple “buy/sell” factors (Baker, Bloom & Davis, 2016).

Our set-up (monthly SOXX volatility; attention + entropy; tree-based model; SHAP) can be used for:

- Risk diagnostics (crowding and fragility overlay): track attention growth together with entropy drops. When attention rises while entropy falls, treat it as a “crowded narrative” condition in which small surprises can produce outsized volatility moves.
- Regime detection: define regimes by the joint state of attention and entropy (high-attention/low-entropy (convergent focus) vs high-attention/high-entropy (contested stories)).
- Thematic monitoring (early warning of theme rotation): use rapid entropy change as a “story re-ordering” indicator: sharp declines suggest fast convergence, while sharp rises can indicate competing interpretations and unstable outcomes.

However, we also mention 2 cautions to follow:

- Do not treat narratives as a stable alpha factor: the average MAE gain is small (around 0.083 to 0.082), and the strongest effects are episodic. A simple “high attention = trade” rule is not supported by our regime story.
- Measurement and small-sample risk are real: we used a relatively small monthly sample. This raises obvious risks of provider-specific bias, taxonomy drift, and overfitting to rare episodes.

Limitations & Research Extensions

Before closing this paper, we have to acknowledge several limitations to this research.

First, the analysis is conducted at a monthly frequency with a final sample of 81 observations, which necessarily constrains statistical power and limits the scope for formal hypothesis testing. While tree-based models are well suited to small samples, the results should be interpreted as indicative rather than definitive.

Second, narrative attention is proxied using online information demand rather than direct measures of investor beliefs. Although attention-based proxies are widely used in the literature and align well with the theoretical notion of narrative salience, they do not capture sentiment polarity or belief conviction directly. Future work could integrate richer textual embeddings or survey-based belief measures.

Third, the focus on a single sector ETF implies that the findings may not generalise mechanically to other asset classes or market structures. However, the semiconductor sector's exposure to innovation cycles and geopolitical risk makes it an especially informative case study for narrative-driven dynamics.

Finally, the machine learning framework is deliberately not optimised for trading performance. The goal is explanatory clarity rather than alpha generation. As a result, the model should not be interpreted as a deployable forecasting system, but as a diagnostic tool for understanding how narratives interact with volatility.

Despite these limitations, the results provide robust evidence that narratives improve volatility forecasting in absolute terms and play a meaningful role in shaping market dynamics under specific attention regimes. By combining narrative economics with explainable machine learning, the paper offers a transparent and empirically grounded framework for analysing how stories, attention, and coordination influence financial risk.

Conclusion

Our paper motivates a broader perspective on sector volatility: volatility is not shaped only by shocks to fundamentals; it is also shaped by periodic shifts in collective attention that synchronise market interpretation. This perspective is consistent with narrative economics, where stories spread and temporarily coordinate behaviour, and with limited-attention models where the market's effective information set is filtered by what investors can process.

Empirically, our ML framework turns this perspective into testable structure. We show that narrative variables deliver incremental forecasting value in absolute-error terms, but more importantly, SHAP diagnostics reveal that narrative attention and entropy primarily act as conditional amplifiers. Indeed, when attention is high and concentrated, predicted volatility increases disproportionately. When attention is low, narrative structure matters far less. In that sense, the key object our analysis identifies is not "narrative level", but "attention synchronisation". Thus, this translates a diffuse information environment into coordinated positioning and fragility.

The resulting quantitative lesson is deliberately restrained: narratives are best treated as state variables for risk monitoring, regime detection, and thematic oversight, complementing rather than replacing macro-financial controls. This framing also travels well beyond semiconductors as a case study: any market segment shaped by salient public discourse, policy uncertainty, or geopolitical exposure may display volatility that depends as much on collective attention shifts as on contemporaneous fundamental shocks.

To finish off, our message is that the real regime shift is not in fundamentals, but in focus. Once many investors lock onto the same story, the market's reaction function steepens, and small shocks can look like structural breaks. The deepest source of risk is not what changes in the world, but when the market starts seeing the world in the same way at the same time.

Appendix & Bibliography

References

Shiller, R. J. (2017). Narrative economics. American Economic Review, 107(4), 967–1004. https://doi.org/10.1257/aer.107.4.967

Shiller, R. J. (2020). Narrative economics: How stories go viral and drive major economic events. Princeton University Press.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. Journal of Finance, 53(1), 307–343. https://doi.org/10.1111/0022-1082.00027

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. Quarterly Journal of Economics, 131(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. Journal of Finance, 62(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66(1). https://doi.org/10.1111/j.1540-6261.2010.01625.x

Mullainathan, S., & Shleifer, A. (2005). The market for news. American Economic Review, 95(4), 1031–1053. https://doi.org/10.1257/0002828054825569

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

If you're interested in exploring the results in more detail, we invite you to our GitHub repository:

https://github.com/EsfsQuantitativeResearch/Narrative-Contagion-in-Markets---ESFS-Quant-Research.git