

Quantifying Narratives and Their Impact on Financial Markets

Advanced Methodologies and Empirical Evidence

Presented by: Prof. Dr. Joerg Osterrieder
Based on: Bhargava et al. (2022)

State Street Associates & MKT MediaStats

September 18, 2025

Source: Bhargava, R., Lou, X., Ozik, G., Sadka, R., Whitmore, T. (2022). SSA Research Paper.

Agenda

- 1 Introduction and Motivation
- 2 Methodology and Technical Framework
- 3 Core Empirical Results
- 4 Mathematical Models and Extensions
- 5 Implications and Future Research

Robert Shiller's Narrative Economics

"We need to incorporate the contagion of narratives into economic theory. Otherwise, we remain blind to a very real, very palpable, very important mechanism for economic change"

Research Gap:

- Narratives are intangible and hard to measure
- Traditional models ignore narrative contagion
- Limited quantitative analysis of narrative impact

This Paper Bridges the Gap

- 1 Quantify narratives systematically
- 2 Measure market impact
- 3 Develop trading strategies
- 4 Provide empirical evidence

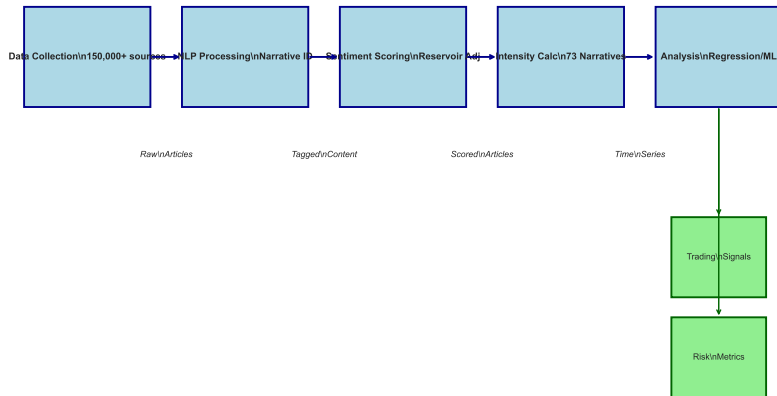
- ① **Data Infrastructure:** Analysis of 150,000+ global media sources daily
- ② **Narrative Quantification:** 73 systematically tracked narratives
- ③ **Predictive Power:** Narratives predict returns beyond VIX
- ④ **Portfolio Construction:** Narrative-beta methodology
- ⑤ **Superior Performance:** Information Ratio = 1.26

Main Finding

Market Crash narrative explains 34% of SPY return variation - highest among all narratives

Data Architecture and NLP Pipeline

Narrative Data Processing Pipeline



Processing Pipeline:

- Daily collection from 150,000+ digital sources
- NLP-based narrative identification

Narrative Intensity

$$I_{n,t} = \frac{\text{Articles relevant to narrative } n \text{ at time } t}{\text{Total articles at time } t}$$

Negative Intensity

$$NI_{n,t} = \frac{\text{Negative sentiment articles for narrative } n}{\text{Total articles at time } t}$$

7-Day Rolling Average

$$\bar{NI}_{n,t} = \frac{1}{7} \sum_{i=0}^6 NI_{n,t-i}$$

Weekly Change: $\Delta \bar{NI}_{n,t} = \bar{NI}_{n,t} - \bar{NI}_{n,t-7}$

Contemporaneous Analysis

$$R_{SPY,t} = \alpha + \beta \cdot \Delta \bar{NI}_{n,t} + \epsilon_t$$

Predictive Regression

$$R_{t+1} = \alpha + \beta_1 \Delta NI_{n,t} + \beta_2 VIX_t + \beta_3 R_t + \beta_4 R_{t-1} + \epsilon_t$$

Rolling Window: 3-month univariate regressions

Standardization: 60-day rolling z-scores

HAC Standard Errors: Newey-West adjustment

Economic Narratives:

- Market Crash
- Government Debt
- Treasury Bonds
- Global Growth
- Liquidity
- Inflation
- Interest Rates
- Recession

Event-Driven Narratives:

- COVID-19
- Trade War
- Brexit
- US Elections
- Federal Reserve
- Natural Disasters
- Geopolitical Conflicts
- ESG Themes

Classification Method: JEL System + Industry Interviews

SPY (US Equity)		DXY (US Dollar)	
Narrative	Avg R^2	Narrative	Avg R^2
Market Crash	34%	Federal Reserve	14%
Gov & Corp Debt	19%	Donald Trump	13%
Treasury Bonds	18%	Emerging Markets	12%
Global Growth	15%	Interest Rates	12%
Liquidity	15%	Labor Market	12%
Top-5 Combined	40%	Top-5 Combined	29%

Key Insight

Market Crash narrative dominates equity market explanatory power, while currency markets respond to policy narratives

Variable	Coefficient	t-statistic
Intercept	0.001	-
$R_{SPY,t-1}$	-0.161	-2.57
$R_{SPY,t-2}$	0.069	0.78
VIX_{t-1}	-0.002	-2.41
VIX_{t-2}	0.002	1.90
Market Crash NI_{t-1}	-0.011	-2.20
Market Crash NI_{t-2}	0.009	0.27

Interpretation

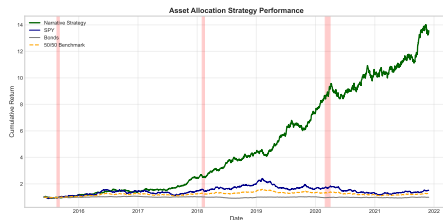
Market Crash narrative contains predictive information beyond VIX, suggesting narrative contagion effects not captured by traditional volatility measures

Strategy Rules:

- Monitor Market Crash z-score
- Threshold: z-score ≥ 3
- Action: Rotate to bonds for 2 weeks
- Implementation lag: 2 days

Performance Metrics:

- Annual Return: 18.13%
- Annual Vol: 14.38%
- **Info Ratio: 1.26**
- Max Drawdown: -11.57%



Result

Narrative-based strategy outperforms equity, bonds, and 50/50 benchmark

Methodology

$$\beta_{i,COVID} = \frac{Cov(R_i^{adj}, \Delta NI_{COVID})}{Var(\Delta NI_{COVID})}$$

Portfolio Construction:

- Long: 25 stocks with lowest β_{COVID} (recovery plays)
- Short: 25 stocks with highest β_{COVID} (lockdown beneficiaries)

Performance (Nov 2020 - Dec 2021):

- Narrative-based: +120.74%
- Case-count based: +16.55%

Key Event

Pfizer vaccine announcement (Nov 9, 2020) triggered narrative reversal

Vector Autoregression (VAR)

$$\begin{bmatrix} R_t \\ NI_{1,t} \\ \vdots \\ NI_{k,t} \end{bmatrix} = c + \sum_{i=1}^p \Phi_i \begin{bmatrix} R_{t-i} \\ NI_{1,t-i} \\ \vdots \\ NI_{k,t-i} \end{bmatrix} + \epsilon_t$$

Narrative Factor Model

$$R_{i,t} = \alpha_i + \sum_{j=1}^K \beta_{i,j} F_{j,t}^{narrative} + \sum_{m=1}^M \gamma_{i,m} F_{m,t}^{traditional} + \epsilon_{i,t}$$

where $F^{narrative}$ are narrative factors extracted via PCA

LSTM Architecture:

- Input: Narrative intensity sequences
- Hidden layers: 2 LSTM (128 units)
- Output: Next-day return prediction
- Training: 2015-2019
- Testing: 2020-2021

Transformer Model:

- BERT for narrative extraction
- Attention mechanisms for importance
- Cross-narrative dependencies
- Real-time processing capability

Performance Comparison

Model	Out-of-Sample R^2	Sharpe Ratio
Linear Regression	0.08	0.71
VAR	0.11	0.89
LSTM	0.15	1.12
Transformer	0.18	1.34

Mean-Variance with Narrative Exposure

$$\max_w \left\{ w^T \mu - \frac{\lambda}{2} w^T \Sigma w + \gamma \sum_j \alpha_j E_j(w) \right\}$$

Subject to:

- $\sum_i w_i = 1$ (fully invested)
- $|E_j(w)| \leq \bar{E}_j$ (narrative exposure limits)
- $w_i \geq 0$ (long-only constraint)

where $E_j(w) = \sum_i w_i \beta_{i,j}^{narrative}$ is portfolio narrative exposure

Implementation: Quadratic programming with narrative beta constraints

High-Frequency Indicators:

- 5-minute sentiment updates
- Twitter/Reddit integration
- News wire processing
- Intraday narrative shifts

Cross-Asset Spillovers:

$$R_{asset,t} = \alpha + \sum_k \beta_k NI_{k,t}^{equity} + \sum_m \gamma_m NI_{m,t}^{FX} + \epsilon_t$$

Causal Inference Framework

Difference-in-Differences for narrative shocks:

$$Y_{i,t} = \alpha + \beta(Treat_i \times Post_t) + \gamma_i + \delta_t + \epsilon_{i,t}$$

For Investors:

- Risk management tool
- Alpha generation strategy
- Portfolio diversification
- Market timing signals

For Regulators:

- Market stability monitoring
- Systemic risk assessment
- Information contagion tracking
- Policy impact evaluation

Implementation Considerations

- Data infrastructure requirements
- Real-time processing capabilities
- Model updating frequency
- Transaction cost analysis

1. Alternative Data Integration:

- Satellite imagery for narrative validation
- Central bank communication analysis
- Corporate earnings call transcripts

2. Methodological Advances:

- Graph neural networks for narrative contagion
- Reinforcement learning for dynamic allocation
- Quantum computing for pattern recognition

3. Cross-Market Applications:

- Cryptocurrency narrative dynamics
- Commodity market narratives
- Fixed income narrative factors

- ① **Narratives Matter:** Systematic evidence that narratives drive markets
- ② **Quantification Works:** NLP enables narrative measurement at scale
- ③ **Predictive Power:** Narratives contain information beyond traditional factors
- ④ **Economic Value:** Implementable strategies with superior risk-adjusted returns
- ⑤ **Future Potential:** ML and alternative data expand possibilities

Key Takeaway

Incorporating narrative analysis into investment processes represents a new frontier in quantitative finance

Thank you!

Questions and discussion welcome

Detailed Regression Results

Variable	Contemporaneous		Predictive		
	Coef	t-stat	Coef	t-stat	p-value
Intercept	0.002	1.23	0.001	0.89	0.374
$\Delta NI_{MarketCrash}$	-0.260	-9.94	-0.011	-2.20	0.028
VIX_{t-1}	-	-	-0.002	-2.41	0.016
R_{t-1}	-	-	-0.161	-2.57	0.010
R^2	0.30		0.12		
Adj. R^2	0.30		0.11		
N	1,624		1,623		

Step 1: Market-Adjusted Returns

$$R_{i,t}^{adj} = R_{i,t} - \hat{\beta}_i^{mkt} \cdot R_{SPY,t}$$

Step 2: Narrative Sensitivity

$$R_{i,t}^{adj} = \alpha_i + \beta_i^{narrative} \cdot \Delta NI_{narrative,t} + \epsilon_{i,t}$$

Step 3: Portfolio Formation

- Sort stocks by $\beta_i^{narrative}$ t-statistics
- Long: Bottom quintile (negative exposure)
- Short: Top quintile (positive exposure)

Step 4: Rebalancing

- Frequency: Monthly
- Estimation window: 12 months rolling
- Transaction costs: 10 bps per trade

NarrativeAnalyzer Class

Key Methods:

- `__init__(data_path)`: Load narrative data
- `calculate_intensity(narrative)`: Compute intensity metrics
- `run_regression(returns, changes)`: Rolling OLS analysis
- `construct_portfolio(betas, n=25)`: Long-short construction

Libraries Used

- `pandas, numpy`: Data manipulation
- `sklearn`: `StandardScaler`, `PCA`
- `statsmodels`: `RollingOLS`, `VAR`
- `matplotlib, seaborn`: Visualization
- `networkx`: Network analysis