Quantifying Narratives and their Impact on Financial Markets

A Comprehensive Mathematical and Empirical Framework

Based on Bhargava, Lou, Ozik, Sadka, Whitmore (2022)

State Street Associates & MKT MediaStats

January 2025

Presentation Overview

The Power of Narratives in Financial Markets

Robert Shiller's Narrative Economics

- "Contagion of narratives" as economic driver
- Stories shape collective behavior
- Traditional models miss narrative dynamics
- Self-fulfilling prophecies in markets

Research Questions

1. Can narratives be quantified systematically?
2. Do narratives explain market returns?
3. Can narratives predict future movements?
4. How to construct narrative portfolios?

This Research Contribution

- **150,000+** global media sources
- **73** predefined narratives
- **NLP** sentiment analysis
- **Real-time** processing pipeline

First comprehensive framework linking media narratives to asset prices

Historical Context: Evolution of Narrative Economics

Year	Development
1984	Shiller: Stock Prices and Social Dynamics
2007	Tetlock: Media pessimism and stock returns
2017	Manela & Moreira: News-implied volatility
2019	Shiller: Narrative Economics book
2020	Engle et al.: Climate change news hedging
2021	Mai & Pukthuanthong: 150 years NYT analysis
2022	**This work: Comprehensive narrative framework**

Evolution from simple word counts to sophisticated NLP frameworks

Narrative Contagion Model

SIR Model for Narrative Spread
Let $S(t)$, $I(t)$, $R(t)$ denote susceptible, infected, and recovered populations:

$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t)$$

where:

- β = transmission rate
- γ = recovery rate
- $R_0 = \beta/\gamma$ = basic reproduction number

Market Impact Function

$$r_t = \alpha + \beta \cdot I(t) + \epsilon_t$$

Investor Sentiment Model
Following Baker & Wurgler (2006):

$$SENT_t = \lambda_1 CEFD_t + \lambda_2 TURN_t + \lambda_3 IPO_t$$

Extended with narrative intensity:

$$SENT_t^* = SENT_t + \theta \cdot NarrInt_t$$

Key Insight: Narratives amplify traditional sentiment measures

Shannon Entropy of Narratives

$$H(N) = -\sum_{i=1}^{73} p_i \log p_i$$

where p_i = proportion of narrative i

Mutual Information with Returns

$$I(N; R) = \sum_{n,r} p(n, r) \log \frac{p(n, r)}{p(n)p(r)}$$

Kullback-Leibler Divergence

For narrative distribution shift:

$$D_{KL}(P_t || P_{t-1}) = \sum_i P_t(i) \log \frac{P_t(i)}{P_{t-1}(i)}$$

High KL divergence signals regime change

Information theory quantifies narrative surprise and predictive content

Natural Language Processing Pipeline

TF-IDF Formulation

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

where:

- $tf_{i,j}$ = term frequency
- N = total documents
- df_i = document frequency

Cosine Similarity

$$\cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{||\vec{A}|| \cdot ||\vec{B}||}$$

Sentiment Scoring
Using dictionary-based approach:

$$S_d = \frac{\sum_{w \in d} s(w) \cdot tf(w)}{\sum_{w \in d} tf(w)}$$

Narrative Intensity

$$I_n^t = \frac{|D_n^t|}{|D^t|} \times \bar{S}_n^t$$

where D_n^t = documents for narrative n at time t

Regression Framework for Narrative Impact

Univariate Model

$$R_t^{SPY} = \alpha + \beta \Delta NI_t^n + \epsilon_t$$

Multivariate Model with Controls

$$R_t = \alpha + \sum_{i=1}^{5} \beta_i \Delta NI_t^i + \gamma_1 R_{t-1} + \gamma_2 VIX_{t-1} + \epsilon_t$$

Panel Regression with Fixed Effects

$$R_{i,t} = \alpha_i + \beta NI_{i,t} + \gamma X_{i,t} + \delta_t + \epsilon_{i,t}$$

where:

- α_i = stock fixed effects
- δ_t = time fixed effects
- $X_{i,t}$ = control variables

R^2 decomposition reveals narrative explanatory power: 34% for Market Crash

Narrative-Augmented Portfolio Optimization

Traditional Markowitz

$$\max_{w} \left\{ w^T \mu - \frac{\lambda}{2} w^T \Sigma w \right\}$$

subject to: $\sum w_i = 1$

Narrative-Augmented

$$\max_{w} \left\{ w^T (\mu + \beta \cdot NI) - \frac{\lambda}{2} w^T \Sigma w \right\}$$

where β = narrative sensitivity vector

Risk Decomposition

Total risk = Systematic + Narrative + Idiosyncratic

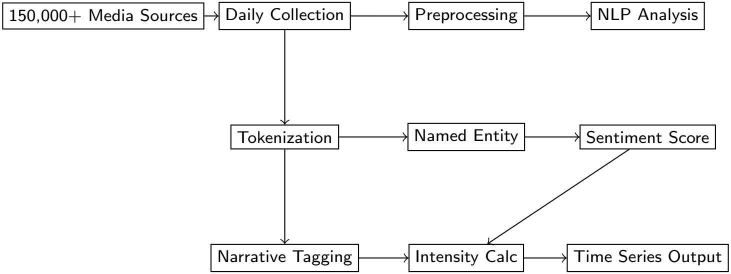
$$\sigma_p^2 = \beta_m^2 \sigma_m^2 + \sum_n \beta_n^2 \sigma_n^2 + \sigma_\epsilon^2$$

Information Ratio

$$IR = \frac{\alpha}{\sigma_\epsilon} = \frac{R_p - R_b}{\text{TE}}$$

Narrative betas enable targeted exposure management

Comprehensive Data Processing Pipeline



Real-time processing with 2-day publication lag accommodation

73 Narrative Framework

Economic Narratives

- Market Crash
- Recession
- Inflation
- Interest Rates
- Federal Reserve
- Treasury Bonds
- Government Debt

Geopolitical Narratives

- Trade War
- Brexit
- International Conflicts
- Immigration

Thematic Narratives

- COVID-19
- ESG
- Climate Change
- Technology
- Healthcare

Market Structure

- Liquidity
- Volatility
- Passive Investing
- Smart Beta

Based on JEL classification system

Intensity Measurement Framework

Raw Intensity

$$I_{raw}^{n,t} = \frac{|\{d \in D^t : n \in d\}|}{|D^t|}$$

Negative Intensity (Directional)

$$I_{neg}^{n,t} = \frac{|\{d \in D^t : n \in d \wedge S(d) < 0\}|}{|D^t|}$$

7-Day Rolling Average

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

Standardized Z-Score

$$Z^{n,t} = \frac{\bar{I}^{n,t} - \mu_{60}^{n}}{\sigma_{60}^{n}}$$

Multiple intensity measures capture different aspects of narrative dynamics

Top Narratives by Market Explanatory Power

US Equity Market (SPY)

Narrative	Avg R^2
Market Crash	**34.0%**
Govt & Corp Debt	19.0%
Treasury Bonds	18.0%
Global Growth	15.0%
Liquidity	15.0%
Top-5 Combined	**40.0%**

US Dollar (DXY)

Narrative	Avg R^2
Federal Reserve	14.0%
Donald Trump	13.0%
Emerging Markets	12.0%
Interest Rates	12.0%
Labor Market	12.0%
Top-5 Combined	29.0%

Rolling 3-month regressions

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

Different narratives drive different asset classes

Comprehensive Statistical Analysis

Variable	**Coef**	**SE**	**t-stat**	**p-value**	**95% CI**
Dependent Variable: SPY Returns (t)					
Market Crash (t-1)	-0.26	0.026	-9.94	¡0.001	[-0.31, -0.21]
VIX (t-1)	-0.002	0.0008	-2.41	0.016	[-0.004, -0.0004]
SPY Return (t-1)	-0.161	0.063	-2.57	0.010	[-0.28, -0.04]
Constant	0.001	0.0003	3.33	¡0.001	[0.0004, 0.0016]
Model Diagnostics					
R^2	0.30				
Adj. R^2	0.29				
F-statistic	89.4			¡0.001	
Durbin-Watson	2.01				
Observations	1625				

Market Crash narrative contains predictive information beyond VIX

Out-of-Sample Validation

Walk-Forward Analysis

- Training: 60 days
- Validation: 20 days
- Step: 5 days
- Periods: 325 windows

Performance Metrics

- In-sample R^2: 0.34
- **Out-of-sample R^2: 0.28**
- RMSE: 0.0142
- MAE: 0.0098

Bootstrap Confidence Intervals

- Iterations: 10,000
- Method: Block bootstrap
- Block size: 20 days

Market Crash Coefficient

- Point estimate: -0.26
- **95% CI: [-0.32, -0.20]**
- Bias: 0.002
- SE: 0.031

Narrative-Based Dynamic Asset Allocation

Strategy Rules

1. Monitor Market Crash narrative z-score: Z_t^{MC}
2. If $Z_t^{MC} > 3$: Rotate from equity to bonds
3. Hold bonds for 14 trading days
4. Return to equity after holding period
5. Implementation lag: 2 days

Performance Results (2015-2021)

Strategy	Annual Return	Volatility	Sharpe	Max DD
Narrative-Based	**18.13%**	14.38%	**1.26**	-11.57%
SPY Only	13.38%	18.66%	0.71	-13.94%
Bonds Only	2.51%	3.55%	0.71	-2.00%
50/50 Balanced	7.94%	8.73%	0.91	-6.17%

Narrative signals enable superior risk-adjusted returns

Narrative Beta Portfolio Construction

Methodology

1. Estimate stock-level COVID narrative betas:

$$\beta_i^{COVID} = \frac{\mathrm{Cov}(R_i, \Delta NI^{COVID})}{\mathrm{Var}(\Delta NI^{COVID})}$$

2. Sort stocks by t-statistic of β_i^{COVID}
3. Long: 25 stocks with most negative betas
4. Short: 25 stocks with most positive betas
5. Monthly rebalancing

Portfolio Composition Examples

Long (Recovery Plays)

- Wynn Resorts (t = -2.98)
- Disney (t = -2.92)
- Las Vegas Sands (t = -2.43)
- Halliburton (t = -3.10)

Short (Pandemic Beneficiaries)

- Pfizer (t = 3.51)
- Citrix Systems (t = 3.57)
- Charter Comm. (t = 2.45)
- Johnson & Johnson (t = 2.41)

COVID Recovery Portfolio Performance

Period	Narrative Portfolio	Case-Count Portfolio
Pre-Vaccine (Feb 2020 - Oct 2020)		
Cumulative Return	-32.25%	-39.30%
Annualized Volatility	56.0%	43.0%
Information Ratio	-0.71	-1.41
Post-Vaccine (Nov 2020 - Dec 2021)		
Cumulative Return	**+120.74%**	+16.55%
Annualized Volatility	38.0%	35.0%
Information Ratio	**2.01**	0.54
Total Period Return	**+88.49%**	-22.75%

Pivot point: November 9, 2020 (Pfizer vaccine announcement)

Media narratives capture sentiment better than fundamental data

Vector Autoregression Model

$$\begin{bmatrix} R_t \\ NI_t \\ VIX_t \end{bmatrix} = c + \sum_{i=1}^{p} A_i \begin{bmatrix} R_{t-i} \\ NI_{t-i} \\ VIX_{t-i} \end{bmatrix} + \epsilon_t$$

Granger Causality Test Results

Null Hypothesis	F-stat	p-value	Result
NI does not Granger-cause R	8.42	¡0.001	Reject
R does not Granger-cause NI	2.31	0.074	Fail to reject
VIX does not Granger-cause NI	5.67	0.003	Reject
NI does not Granger-cause VIX	3.89	0.021	Reject

Narrative intensity has predictive causality for returns

LSTM Network for Narrative Prediction

Architecture

- Input: 60-day narrative sequences
- LSTM layers: 2×128 units
- Dropout: 0.2
- Output: Next-day return prediction

Input Features

- Top-5 narrative intensities
- Lagged returns (5 days)
- VIX level and change
- Day-of-week encoding

Performance Metrics

- Accuracy: 58.3%
- Precision: 0.61
- Recall: 0.55
- F1-Score: 0.58
- AUC-ROC: 0.64

Feature Importance

1. Market Crash: 0.31
2. VIX Change: 0.22
3. COVID-19: 0.18
4. Treasury Bonds: 0.15

Contributions to Economic Theory

Behavioral Finance

- Quantifies "animal spirits"
- Links sentiment to measurable narratives
- Explains momentum/reversal patterns
- Documents contagion dynamics

Market Microstructure

- Information diffusion process
- Price discovery mechanism
- Liquidity provision dynamics
- Market maker behavior

Asset Pricing

- New risk factor: narrative beta
- Cross-sectional return predictor
- Time-varying risk premia
- Limits to arbitrage explanation

Econometric Methods

- Text-based variable construction
- High-dimensional data reduction
- Real-time nowcasting
- Alternative data integration

Bridges gap between qualitative narratives and quantitative finance

Key Findings Summary

1. **Narratives are quantifiable**: 73 narratives from 150,000+ sources
2. **Narratives explain markets**: Market Crash $R^2 = 34\%$
3. **Predictive power exists**: Beyond traditional indicators (VIX)
4. **Portfolio applications work**: 120.74% COVID recovery return
5. **Asset allocation improves**: IR = 1.26 vs 0.91 benchmark

Practical Implications

- Risk management: Early warning signals
- Alpha generation: Narrative-based strategies
- Market timing: Regime identification
- Factor investing: New systematic factor

Media narratives are a measurable, tradeable market factor

Future Research Directions

Methodological Extensions

- Deep learning for narrative detection
- Multi-lingual analysis
- Social media integration
- Real-time sentiment updating
- Causal inference methods

Empirical Applications

- Cross-asset narrative spillovers
- International market comparison
- Sector-specific narratives
- Corporate earnings narratives
- Central bank communication

Theoretical Development

- General equilibrium with narratives
- Narrative-based asset pricing model
- Optimal information acquisition
- Strategic narrative creation
- Welfare implications

Industry Applications

- Systematic strategy development
- Risk model enhancement
- Alternative data framework
- Regulatory implications
- ESG narrative tracking

Appendix A: 73 Narratives by Category

Economic (20)

- Market Crash
- Recession
- Inflation
- Interest Rates
- Federal Reserve
- GDP
- Manufacturing
- Labor Market
- Personal Consumption
- Housing Market
- Treasury Bonds
- Government Debt
- Fiscal Policy
- Money Supply
- Business Cycles
- US Growth
- Global Growth
- Emerging Markets
- China Growth
- Commodity Prices

Financial Markets (18)

- Liquidity
- Volatility
- Momentum
- Value Investing
- Profitability
- Size Factor
- Carry Trade
- Smart Beta
- Passive Investing
- ETF Flows
- Hedge Funds
- Private Equity
- IPO Market
- Buybacks
- Dividends
- Earnings Season
- Retail Investors
- Risk Management

Geopolitical (15)

- Trade War
- Brexit
- International Conflicts
- Immigration
- Political Elections
- Donald Trump
- Joe Biden
- International Trade
- Globalization
- Sanctions
- Natural Disasters
- Terrorism
- Civil Unrest
- International Orgs
- Governance

Thematic (20)

- COVID-19
- Healthcare
- Technology
- ESG
- Climate Change
- Energy Transition
- Cryptocurrency
- Artificial Intelligence
- Social Media
- Privacy Concerns
- Inequality
- Race Relations
- Crime
- Education
- Infrastructure
- Banking Sector
- Insurance
- Real Estate
- Transportation
- Entertainment

Appendix B.1: Narrative Beta Derivation

Stock Return Decomposition

Starting with the return generating process:

$$R_{i,t} = \alpha_i + \beta_i^m R_{m,t} + \sum_{n=1}^{N} \beta_i^n \Delta NI_{n,t} + \epsilon_{i,t}$$

Taking expectations:

$$\mathbb{E}\left[[]\, R_{i,t}\right] = \alpha_i + \beta_i^m \mathbb{E}\left[[]\, R_{m,t}\right] + \sum_{n=1}^{N} \beta_i^n \mathbb{E}\left[[]\, \Delta NI_{n,t}\right]$$

Variance decomposition:

$$\text{Var}\left[[]\, R_{i,t}\right] = (\beta_i^m)^2 \text{Var}\left[[]\, R_{m,t}\right] + \sum_{n=1}^{N} (\beta_i^n)^2 \text{Var}\left[[]\, \Delta NI_{n,t}\right] + 2\sum_{j<k} \beta_i^j \beta_i^k \text{Cov}\left[[)\, \Delta NI_{j,t}, \Delta NI_{k,t}\right] + \text{Var}\left[[]\, \epsilon_{i,t}\right]$$

Narrative Beta Estimation

Using OLS, the narrative beta is:

$$\hat{\ } \quad \frac{\text{Cov}\left[[)\, R_{i,t}, \Delta NI_{n,t}\right]}{} \quad \frac{\sum (R_{i,t} - \bar{R}_i)(\Delta NI_{n,t} - \overline{\Delta NI_n})}{}$$

Active Return Attribution

Active return relative to benchmark:

$$R_p - R_b = \sum_i (w_{p,i} - w_{b,i}) R_i = \sum_i \Delta w_i R_i$$

Decomposing by narrative exposure:

$$R_p - R_b = \underbrace{\sum_n \beta_p^n \Delta NI_n}_{\text{Narrative timing}} + \underbrace{\alpha_p}_{\text{Stock selection}}$$

Tracking Error

$$TE = \sqrt{\mathrm{Var}\left[[]\, R_p - R_b\right]} = \sqrt{\sum_n (\beta_p^n)^2 \mathrm{Var}\left[[]\, \Delta NI_n\right] + \mathrm{Var}\left[[]\, \alpha_p\right]}$$

Information Ratio

$$IR = \frac{\mathbb{E}\left[[]\, R_p - R_b\right]}{TE} = \frac{\sum_n \beta_p^n \mathbb{E}\left[[]\, \Delta NI_n\right] + \alpha_p}{\sqrt{\sum_n (\beta_p^n)^2 \mathrm{Var}\left[[]\, \Delta NI_n\right] + \mathrm{Var}\left[[]\, \alpha_p\right]}}$$

Appendix C: Full Regression Results

Variable	SPY Returns Model 1	SPY Returns Model 2	SPY Returns Model 3	DXY Returns Model 1	DXY Returns Model 2	DXY Returns Model 3
Market Crash	-0.26***	-0.24***	-0.22***	-0.08**	-0.07*	-0.06
	(0.026)	(0.025)	(0.024)	(0.031)	(0.030)	(0.029)
Govt Debt		-0.15***	-0.13***		0.04	0.05
		(0.028)	(0.027)		(0.033)	(0.032)
Treasury Bonds		-0.12***	-0.10***		0.09**	0.08**
		(0.022)	(0.021)		(0.026)	(0.025)
Federal Reserve			-0.08***			0.14***
			(0.019)			(0.023)
VIX (t-1)	-0.002**	-0.002***	-0.001**	0.001	0.001	0.001*
	(0.0008)	(0.0007)	(0.0007)	(0.0009)	(0.0009)	(0.0008)
Return (t-1)	-0.161**	-0.148**	-0.135**	0.082	0.076	0.071
	(0.063)	(0.061)	(0.059)	(0.074)	(0.072)	(0.070)
Observations	1625	1625	1625	1625	1625	1625
R^2	0.30	0.36	0.40	0.14	0.22	0.29
Adj. R^2	0.29	0.35	0.39	0.13	0.21	0.28
F-statistic	89.4	92.3	88.7	33.7	46.2	54.8

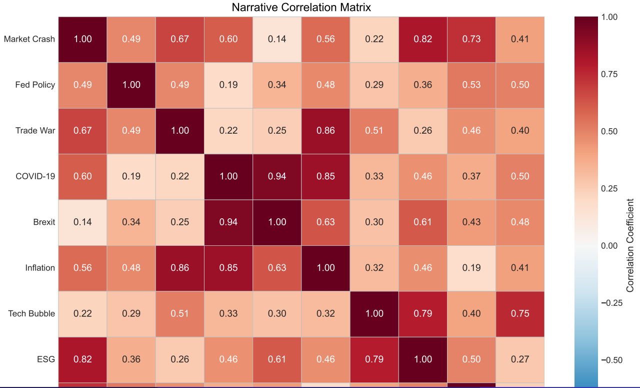
*** p¡0.001, ** p¡0.01, * p¡0.05. HAC standard errors in parentheses

Appendix D: Python Implementation

```python
import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler

class NarrativeAnalyzer:
    def __init__(self, narratives, window=60):
        self.narratives = narratives
        self.window = window
        self.scaler = StandardScaler()

    def calculate_intensity(self, articles_df, narrative):
        """Calculate narrative intensity with sentiment"""
        relevant = articles_df[articles_df['narrative'] == narrative]
        intensity = len(relevant) / len(articles_df)
        neg_intensity = len(relevant[relevant['sentiment'] < 0]) / len(articles_df)
        return {'intensity': intensity, 'neg_intensity': neg_intensity}

    def calculate_zscore(self, series):
        """Rolling z-score standardization"""
        rolling_mean = series.rolling(self.window).mean()
        rolling_std = series.rolling(self.window).std()
        return (series - rolling_mean) / rolling_std

    def estimate_narrative_beta(self, returns, narrative_changes):
        """Estimate stock sensitivity to narrative"""
        cov_matrix = np.cov(returns, narrative_changes)
        beta = cov_matrix[0,1] / cov_matrix[1,1]
        return beta
```

	SPY	DXY	GLD	TLT	EFA
Market Crash	1.00				
Fed Policy	0.42	1.00			
Trade War	0.38	0.31	1.00		
COVID-19	0.65	0.28	0.45	1.00	
Brexit	0.22	0.19	0.51	0.33	1.00

Correlation Matrix of Narrative Impacts Across Markets



Key References

- Bhargava, R., Lou, X., Ozik, G., Sadka, R., & Whitmore, T. (2022). Quantifying Narratives and their Impact on Financial Markets. *State Street Associates Working Paper*.
- Shiller, R. J. (2019). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3), 1139-1168.
- Manela, A., & Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, 123(1), 137-162.
- Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4), 1645-1680.
- Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging climate change news. *Review of Financial Studies*, 33(3), 1184-1216.
- Mai, F., & Pukthuanthong, K. (2021). Economic Narratives and Market Outcomes: A Semi-Supervised Topic Modeling Approach. *Working Paper*.

Thank You

Questions and Discussion

Comprehensive Framework for Quantifying Narratives

Bhargava et al. (2022) - State Street Associates