Shifting Stories, Shifting Returns

Komal Niraula

New York University

https://github.com/komalniraula/Shifting-Stories--Shifting-Returns.git

Summary

This study investigates how longitudinal changes in corporate narratives influence subsequent stock returns, highlighting their relevance as a complementary source of predictive information beyond traditional quantitative factors. Building on the growing field of financial text analysis and narrative disclosure, the research leverages quarterly financial reports and earnings call transcripts from the Orbit dataset to construct a comprehensive feature set capturing sentiment, financial and business content, risk disclosure, subjectivity, and temporal framing.

Multiple Moving Target Score (MTS) variants were developed to quantify quarter-over-quarter shifts across these features. Smoother and weighted formulations (rolling, EWMA, median, variance, frequency) produced unstable or negligible returns, while machine-learning methods (PCA, Lasso) were inconsistent. In contrast, distance-based measures such as KL divergence and cosine similarity yielded the most robust performance, achieving Sharpe ratios up to 0.54 with 75% hit rates. Factor regressions further revealed that while beta exposures to market, value, and momentum are often substantial, positive alpha arises when narrative signals capture information not already explained by these standard risk factors. A composite strategy using principal component analysis (PCA) to blend cosine and variance-based signals achieved improved robustness and cumulative returns.

1 Relevance

Textual data have increasingly been used to predict company performance and stock returns. The analysis of market sentiment, management tone, and disclosure statements has shown consistent forecasting power across settings. Understanding the narratives in financial communications has become increasingly important in both research and practice. Unlike purely quantitative metrics, these narratives can reveal hard-to-quantify aspects of firm fundamentals that other disclosures miss. The growing field of financial narrative processing underscores this importance, with surging interest in applying NLP to corporate reports and using state-of-the-art methods (e.g., word embeddings and transformer-based entailment models) to extract insights from qualitative text.

A particularly relevant insight is that changes in a company's narrative over time can be even more informative than static text. When management's tone or emphasis shifts from one quarter to the next, it often signals a change in underlying outlook or conditions. Studies have

shown that tone changes carry significant predictive power. A quarter-over-quarter increase in negativity in earnings calls, termed a "bleak tone change," is regarded as a strong signal of lower future earnings and heightened uncertainty for the firm. Notably, the market tends to initially underreact to these narrative shifts: stock prices move in the expected direction but persistently, indicating that the full implications are not immediately priced in (**druz2020**). This underreaction suggests an opportunity for investors to benefit by paying attention to narrative changes. Indeed, broader evidence shows that shifts in narratives can be exploited: one recent study constructed "narrative momentum" portfolios and found that stocks associated with rising narrative themes outperformed those with declining themes by about 8% annually, delivering alpha beyond standard risk factors (**ssrn2024**).

The "Moving Targets" paper from Harvard (**cohen2020**) adds a critical perspective. The study highlights that managers often alter their KPI framing in response to weakening fundamentals, and that these narrative substitutions are not random. This provides an academic foundation for using longitudinal textual analysis as a leading indicator of risk. All these findings highlight that when corporate storytelling changes, it often foreshadows real outcomes and market movements that systematic analysis can capture.

This work builds on these insights by identifying narrative shifts in quarterly reports and quantifying them for use in a trading strategy. Based on previous research, multiple dimensions of the narrative are examined to capture a comprehensive picture of how it evolves. In particular, this research incorporates:

- Tone (sentiment)

- Financial and business content relevance

- Risk-related disclosures

- Analyst mentions and discussion

- Subjective vs. objective phrasing

- Forward-looking vs. historical orientation

By combining these features, the approach provides a rich, multidimensional view of narrative change. This is highly relevant in today's markets because it merges financial domain knowledge with cutting-edge NLP techniques, enabling systematic detection of subtle shifts in managerial emphasis. As shown later in the methodology and findings sections, such shifts can be transformed into a tradable signal, offering predictive power above and beyond standard factor models.

2 Methodology

The methodology consisted of three main stages: data preparation, feature construction, and signal development with benchmarking and robustness checks.

2.1 Data Preparation

Corporate filings were obtained from the Orbit dataset, which provides quarterly reports (10-K/10-Q MDA sections and earnings call transcripts) with parsed `blocks` and `pages` files. Each report was linked to factor data and aligned with historical stock returns from CRSP.

Positions were entered on the first trading day of the month following report publication to allow the market time to process the disclosure. Positions were exited on the first trading day of the month after the subsequent quarterly report. In robustness checks, alternative exit rules were considered.

2.2 Feature Construction

Narratives were transformed into structured features across multiple categories:

- **Financial:** Sales, Expense, Profit, Operations, Liquidity, Investment, Financing, Litigation, Employment, Tax, Accounting (classified with embeddings, cosine threshold > 0.35).

- **Business:** Product, Buyer, Process, Location, Promotion, Supplier, Management, Strategy, Industry, Business Description (classified with embeddings).

- **Risk:** Firm-specific vs. Generic boilerplate (classified with MNLI).

- **Analyst:** Detected if a sentence originated from an analyst or from management (MNLI).

- **Sentiment:** Negative, Neutral, Positive (FinBERT).

- **Subjective vs. Objective:** Rule-based using indicative words and numerical evidence.

- **Forward vs. Historical:** Rule-based using tense markers and modal verbs.

This produced a "narrative profile" for each firm-quarter, capturing financial and non-financial aspects. The differences between report features were calculated to measure shifts.

2.3 Moving Target Score (MTS)

Beyond variance- or frequency-weighted shifts, we experimented with distance-based and control theory based formulations to calculate moving targets. Among several candidates, two measures stood out for producing the most consistent signals: KL divergence and cosine distance. KL divergence penalizes changes in distributional weight across categories, while cosine distance emphasizes directional similarity in the feature space. The superior performance of these measures may stem from their ability to filter out noisy absolute magnitudes and instead isolate relative structural changes in firm communication.

$$MTS_{KL}(t) = \sum_{i=1}^{N} p_{i,t} \log \left(\frac{p_{i,t}}{p_{i,t-1}} \right), \tag{1}$$

where $p_{i,t}$ and $p_{i,t-1}$ are normalized feature weights in quarters t and $t-1$, respectively.

$$MTS_{\cos}(t) = 1 - \frac{\sum_{i=1}^{N} f_{i,t} f_{i,t-1}}{\sqrt{\sum_{i=1}^{N} f_{i,t}^2} \sqrt{\sum_{i=1}^{N} f_{i,t-1}^2}}, \tag{2}$$

Empirically, MTS_{KL} delivered a Sharpe ratio of 0.46 with 75% hit rate, while $MTS_{\cos}$ achieved an even stronger Sharpe of 0.54 with the same hit rate. Their relative success suggests that distance-based approaches, by capturing proportional and directional narrative shifts rather than absolute changes, provide a more robust mapping from text to returns.

2.4 Factor Analysis

To test whether the predictive power of the Moving Target Scores (MTS) was due to true informational content rather than exposure to standard risk factors, a factor regression analysis was conducted. Specifically, we regressed the MTS-based long-short portfolio returns against a set of factors. The regressions were run using ordinary least squares (OLS) with Newey-West HAC standard errors (1-lag), and alpha significance was evaluated using quarterly returns annualized by a factor of 4.

3 Data Insights

Several observations emerged from the dataset analysis:

- **Bang-Bang Max Delta:** Using the largest quarterly narrative shift variable as a predictor achieved 54.0% accuracy on up returns and 51.0% on down returns, suggesting extreme changes carry the most actionable information.

- **Distribution of Narrative Shifts:** On average, each firm's quarterly report exhibited changes in about 20 narrative features, far higher than the "2–3 shifts per quarter" intuition. This shows that narratives are highly dynamic from quarter to quarter, with many granular adjustments being captured in the deltas.

- **Financial vs. Non-Financial Shifts:** Nearly 9 out of 10 firm-quarters included at least one significant change in financial categories (e.g., Sales, Profit, Liquidity) and at least one change in business (non-financial) categories (e.g., Product, Management, Strategy). This indicates that both financial and non-financial framing are frequently adjusted, and shifts in non-financial narratives often preceded measurable financial deterioration, echoing patterns documented in the "Moving Targets" literature.

- **Risk Language:** Firm-specific risk mentions were sparse, while generic risk disclosures showed the strongest link to returns, with `delta_RiskGeneric_avg` correlated ≈ 0.14 with subsequent performance.

- **Timeframe Orientation:** Firms with above-median forward orientation in their disclosure earned an average return of 30.4% versus 7.1% for firms with lower forward orientation. This suggests a strong focus on future prospects is associated with substantially higher subsequent performance.

- **Sentence Structure (Subjective vs. Objective):** Firms in the top quartile of subjective language averaged a subsequent return of 50.8% compared with only 9.4% for the most objective firms. This suggests that opinion-heavy, narrative-driven disclosures were strongly associated with higher post-report performance in the post-2023 sample.

- **Analyst Interaction:** Sentences classified as analyst-originated accounted for only about 1.6% of the dataset, reflecting that LLMs are not very effective at detecting analyst questions.

- **Strategy Performance:** MTS variants showed mixed results. Baseline smoothers (rolling, EWMA, median) produced large negative returns, while weighted and ML approaches (variance, frequency, PCA, Lasso) offered little alpha. Distance-based measures—KL divergence and Cosine similarity—were more promising (Sharpe up to 0.54, hit rate 75%), though still factor-loaded and statistically weak.

- **Combined Strategy (MTS_combo):** To stabilize performance, a composite signal was constructed by combining the cosine similarity (MTS_cosine) and variance-weighted (MTS_w_var) approaches using a principal component analysis (PCA) method. The resulting portfolio delivered a Sharpe ratio of 0.48 with a cumulative return of 0.58 over eight quarters. This improvement suggests that PCA effectively extracts the common stable component across heterogeneous signals, dampening noise from individual predictors. By mixing a similarity-based measure (cosine) with a dispersion-sensitive weighting scheme (variance), the strategy balances directional strength with robustness, making it more resilient than either input alone.

- **Factor Exposures:** Factor regressions show that moving-average signals (rolling, EWMA, median) delivered large negative spreads with seemingly high alphas, but these were almost entirely explained by extreme loadings on momentum, value, and size. In contrast, variance- and frequency-weighted signals exhibited much weaker betas and negligible alpha, indicating closer alignment with standard factor returns. Distance-based signals such as KL divergence and Cosine produced more differentiated exposures, suggesting they capture narrative-driven structure not spanned by traditional factors, even though their alphas lacked statistical significance.

4 Conclusion

Data analysis revealed that narrative changes are frequent and multidimensional. Nearly every firm–quarter exhibited both financial and non-financial shifts, with risk framing and forward-looking orientation emerging as particularly influential. Extreme shifts (bang-bang deltas) were especially predictive, underscoring that large narrative adjustments carry the greatest informational value.

The exploration of multiple Moving Target Score (MTS) formulations revealed that while smoothing- and weighting-based signals (rolling, EWMA, variance, frequency) produced unstable or negligible returns, distance-based measures such as KL divergence and cosine similarity offered more consistent predictive content. These measures filtered out noise in absolute magnitudes and instead highlighted proportional and directional narrative re-alignments that markets tend to underreact to.

While LLM-based classification captured broad narrative dimensions effectively, it was less capable of reliably detecting analyst-originated statements or subtle dialogue nuances in transcripts. This suggests that hybrid approaches, blending rule-based heuristics and advanced models, are needed to better capture fine-grained interactions.