CURIOSITYBITS LIVING LAB BOSTON · STUDY #01 RESEARCH ARTICLE (V4.3 — VERBATIM SEC EDGAR 8-K/10-Q PANEL)

When Executives Blame the Macroeconomy: Causal Attribution Asymmetry, Timestamp-Aligned Announcement Returns, and the Limits of Post-Earnings Text Predictability

CuriosityBits Living Lab Boston

Abstract

Corporate earnings releases pair mandatory accounting figures with executive narratives that frame operational performance. Grounded in attribution theory, corporate impression management, and financial textual analysis, we test whether *Executive Blame Attribution*—attributing operational shortfalls asymmetrically to external macroeconomic, foreign-exchange, or supply-chain headwinds rather than internal execution—and *Written Guidance Vagueness* explain announcement-window and 21-trading-day post-announcement returns. Analyzing a stratified panel of 180 verbatim Securities and Exchange Commission (SEC) Form 8-K (N = 176) and Form 10-Q (N = 4) filings across 96 S&P 500 corporations (2022–2026), we combine exact Eastern Time acceptance timestamps ($89 pre-open or intraday versus 91 after-hours releases), XBRL Standardized Unexpected Earnings, revenue and earnings growth, and masked narrative scores across three large language models and a deterministic linguistic scorer. Controlling for quantitative surprises, pre-event momentum and volatility, finance-specific lexical tone, and sector and year fixed effects, *Executive Blame Attribution* exhibits a slope of -5.9921% (p = 0.0284) on timestamp-aligned reaction abnormal returns, -5.7166% (p = 0.0016) on filing-date returns, and -11.3201% (p = 0.0002) among pre-open and intraday releases. Over the 21-trading-day post-reaction window, *Written Guidance Vagueness* is statistically null (-4.4713%, p = 0.4956), and timestamp alignment eliminates an apparent filing-date return reversal (-0.3596, p = 0.0232 unaligned versus -0.1297, p = 0.2102 aligned). Out of sample, a constant zero forecast (8.6554% RMSE) outperforms ordinary least squares (9.1165%) and nested ridge regression (8.9149%$).

1. Introduction

When publicly traded corporations disclose quarterly financial results, accounting numbers rarely travel alone. Executives accompany income statements and balance sheets with narrative commentary that interprets operational trends, contextualizes unexpected shortfalls, and frames forward-looking expectations (Davis, Piger, & Sedor, 2011; Henry, 2008). Across corporate reporting settings, managerial prose reflects a tension between two economic motives (Merkl-Davies & Brennan, 2011). Under an informational signaling view,

managers use qualitative commentary to convey private assessments of earnings persistence beyond what summary accounting metrics reveal (Baginski, Hassell, & Kimbrough, 2004; Davis et al., 2011). Under an opportunistic impression management view, managers facing ex post accountability deploy self-serving causal attributions—crediting internal strategy for successes while blaming exogenous macroeconomic, currency, or supply-chain headwinds for disappointing execution—to soften market penalties (Chen, Han, & Tan, 2016; Kimbrough & Wang, 2013).

Whether equity investors accept defensive external blame, penalize it immediately upon disclosure, or underreact to narrative ambiguity over subsequent weeks remains a central question in accounting and financial economics. Foundational textual studies establish that domain-specific financial word tone (Loughran & McDonald, 2015, 2020), linguistic concreteness (Pan, McNamara, Lee, Haleblian, & Devers, 2017), and structural filing modifications (Cohen, Malloy, & Nguyen, 2020) carry valuation-relevant signals. Yet moving from document-level sentiment to targeted rhetorical constructs—such as asymmetric external blame and guidance vagueness—exposes three methodological hazards that routinely distort event studies in modern electronic markets.

First, textual constructs must align with the genre of the underlying corpus (Bae, Hung, & van Lent, 2023; Bochkay, Brown, Leone, & Tucker, 2022). Unlike interactive quarterly conference calls where analysts pose live questions and executives can dodge inquiries, written SEC Form 8-K earnings releases contain unilateral management prose and financial tables without an interactive exchange. Studying written earnings filings therefore requires operationalizing *Written Guidance Vagueness*—omitting concrete numeric outlook ranges in favor of hedged qualitative commentary—and *Executive Blame Attribution* directly as features of written managerial prose while controlling for structured accounting surprises.

Second, measuring announcement and post-announcement returns requires exact alignment between regulatory acceptance timestamps and equity market trading hours. Roughly half of large-capitalization earnings releases are accepted by the SEC Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system at or after the 16:00:00 Eastern Time (ET) market close. Filings accepted between 16:00:00 ET and 17:30:00 ET receive an official filing date t even though regular trading on day t has already closed, whereas filings accepted after 17:30:00 ET roll forward to the next business day. Anchoring return windows to official calendar filing dates rather than exact acceptance timestamps creates two distortions: it contaminates filing-date announcement regressions for after-hours filers with pre-release daytime trading noise, and it embeds the immediate t + 1 post-release reaction inside the putative post-announcement drift window, generating illusory return reversals.

Third, finance and statistical scholarship underscores a sharp divide between in-sample explanatory significance and out-of-sample return predictability (Goyal, Welch, & Zafirov, 2024; Shmueli, 2010; Xu & Liu, 2024). Even when a narrative construct exhibits a robust conditional association with contemporaneous announcement returns, forecasting multi-week post-event returns in liquid mega-capitalization equities confronts severe signal-to-noise limits. Without cross-validation that isolates firms and calendar cohorts across folds—and without benchmarking against a parameter-free constant zero abnormal return forecast—apparent

predictive gains reflect estimation variance and overfitting rather than tradable alpha (Campbell & Thompson, 2007; Goyal et al., 2024).

We address all three challenges across a stratified quota panel of N = 180 verbatim SEC Form 8-K Item 2.02 press releases (N = 176) and Form 10-Q filings (N = 4) issued by 96 S&P 500 firms across all 10 Global Industry Classification Standard (GICS) sectors from 2022 through 2026 (frozen in October 2026, nearly two months after the final post-event window closed). We pair exact Eastern Time EDGAR acceptance timestamps ($89 pre-open or intraday filings versus 91 after-hours filings) with public SEC XBRL accounting surprises—*XBRL Standardized Unexpected Earnings*, *XBRL Year-over-Year Diluted EPS Growth*, and *XBRL Year-over-Year Revenue Growth*—alongside pre-event stock momentum, pre-event volatility, and finance-specific lexical tone. Masked lead-narrative excerpts (≤ 2,500$ characters) are scored independently by three instruction-tuned large language models and a deterministic linguistic scorer, enabling explicit single-rater and multi-rater intraclass correlation diagnostics.

Three empirical findings emerge. First, controlling for XBRL surprises, pre-event return and volatility, finance-specific lexical tone, and sector and year fixed effects, *Executive Blame Attribution* carries a negative slope of -5.9921% (p = 0.0284) on *Timestamp-Aligned Reaction Abnormal Return*, -5.7166% (p = 0.0016) on *Filing-Date Stock Return* (-5.6197%, p = 0.0063 on Form 8-K filings), and -11.3201% (p = 0.0002) among the $89 pre-open and intraday filings. Rather than insulating share prices through external excuses, heavy defensive blame coincides with sharper immediate discounting. Second, over the 21-trading-day post-reaction window, *Written Guidance Vagueness* is statistically null (-4.4713%, p = 0.4956 aligned; -6.5194%, p = 0.3299 unaligned) within cluster-adjusted 80%-power minimum detectable effect bounds of 2.02 to 2.31 percentage points per standard deviation, while aligning post-event windows to the true reaction close eliminates an apparent filing-date return reversal (-0.3596, p = 0.0232 unaligned versus -0.1297, p = 0.2102 aligned). Third, in five-fold leave-one-year-out and disjoint-firm grouped cross-validation, a constant zero abnormal return forecast (8.6554% RMSE) beats unregularized ordinary least squares (9.1165%), scalar shrinkage (8.7464%), and nested ridge cross-validation (8.9149%$).

These results contribute to accounting disclosure research on causal attributions (Baginski et al., 2004; Chen et al., 2016; Kimbrough & Wang, 2013; Koo, Wu, & Yeung, 2012), behavioral attention and event-window methodology (DellaVigna & Pollet, 2009; Hirshleifer, Lim, & Teoh, 2009), and out-of-sample asset-pricing evaluation (Goyal et al., 2024; Loughran & McDonald, 2020; Shmueli, 2010), showing that immediate skepticism toward executive blame coexists naturally with efficient multi-week post-announcement pricing in S&P 500 equities.

2. Theoretical Framework and Hypotheses

2.1 Attribution Theory, Impression Management, and Announcement-Window Pricing

Corporate narrative reporting unfolds under ex post accountability in which managers anticipate how capital markets evaluate operational outcomes (Merkl-Davies & Brennan, 2011). Synthesizing psychological and economic perspectives, Merkl-Davies and Brennan (2011) distinguish discretionary reporting bias, symbolic framing, and self-serving causal attribution. In quarterly earnings press releases—a hybrid genre serving both informational and promotional functions (Henry, 2008)—executives routinely link financial performance to internal strategic execution or external environmental forces.

Attribution theory predicts an asymmetric self-serving pattern: managers credit internal capability for strong results (enhancing attributions) while blaming macroeconomic slowdowns, currency fluctuations, or supply-chain bottlenecks for shortfalls (defensive attributions). Archival evidence confirms that attributions carry valuation consequences beyond quantitative news. Baginski et al. (2004) show across management earnings forecasts that causal attributions occur more frequently alongside bad news and elicit stronger absolute and more negative price reactions holding forecast news constant. Market participants do not accept defensive framing passively: Kimbrough and Wang (2013) demonstrate that investors test the plausibility of self-serving press-release attributions against industry peer announcements and earnings commonality, penalizing firms when peer performance fails to corroborate claimed external shocks. Likewise, Koo et al. (2012) show that textual attributions separating industry-wide shocks from firm-specific actions drive systematic intra-industry information transfers.

Experimental accounting evidence clarifies the mechanism behind investor skepticism. Chen et al. (2016) find that when managers attribute negative guidance news to external factors rather than internal, controllable causes, investors form lower earnings expectations; whereas internal controllable explanations signal corrective agency, external blame signals persistent exposure to uncontrollable headwinds or evasive stewardship. Separating this rhetorical channel from baseline sentiment requires rigorous controls. Davis et al. (2011) establish across roughly 23,000 earnings press releases that net optimistic language predicts future return on assets and short-window announcement returns beyond earnings surprises. Simultaneously, Loughran and McDonald (2015, 2020) demonstrate that general-purpose dictionaries misclassify over 70% to 83% of financial terms, necessitating finance-specific lexical tone controls alongside quantitative accounting surprises (Bochkay et al., 2022). In liquid S&P 500 equities, if asymmetric external blame signals unhedged operational vulnerability or low credibility, share prices should adjust downward immediately upon disclosure:

> ***Hypothesis 1 (Contemporaneous Discounting of Executive Blame Attribution):*** *Executive Blame Attribution in verbatim earnings filings is negatively associated with announcement-window stock returns after controlling for XBRL earnings and revenue surprises, pre-event return and volatility dynamics, finance-specific lexical tone, and sector and year fixed effects.*

2.2 Disclosure Timing, Investor Attention, and Timestamp-Aligned Price Discovery

Evaluating Hypothesis 1 requires accounting for both behavioral attention constraints and institutional trading mechanics. Behavioral finance demonstrates that limited attention shapes the speed of price discovery: DellaVigna and Pollet (2009) document that Friday earnings releases experience a 15% weaker immediate price response, 8% lower abnormal volume, and 70% stronger delayed drift, while Hirshleifer et al. (2009) show that competing same-day announcements dampen immediate reactions and amplify post-announcement drift. Models of investor overconfidence and outcome-dependent self-attribution (Chou, Li, Yin, & Zhao, 2020; Daniel, Hirshleifer, & Subrahmanyam, 1998), alongside evidence on incomplete belief updating and earnings autocorrelation (Abarbanell & Bernard, 1992; Fink, Palan, & Theissen, 2023), underscore that price formation depends critically on when public signals enter active trading sessions.

In U.S. equity markets, the 16:00:00 ET regular-hours close creates a sharp mechanical boundary. When an issuer files before 16:00:00 ET, the official SEC filing date t matches the first trading session in which investors trade on the text ($t_{react} = t$). When an issuer files between 16:00:00 ET and 17:30:00 ET, EDGAR assigns calendar date t as the official filing date even though regular trading on day t closed prior to dissemination; the first post-release trading session is $t_{react} = t + 1$. Consequently, regressing unaligned day-t filing returns on narrative scores mixes post-release price discovery for pre-open filers with pre-release noise for after-hours filers, attenuating after-hours slopes until timestamps are aligned. Moreover, anchoring a 21-trading-day post-event window to the official filing date ($t + 1 \dots t + 21$) embeds the $t + 1$ announcement reaction inside the dependent variable for after-hours filers, manufacturing a spurious return reversal that should vanish once post-event windows begin strictly after the reaction close ($t_{react} + 1 \dots t_{react} + 21$):

> ***Hypothesis 2a (Timing-Regime Attenuation and Alignment Convergence):*** *In unaligned filing-date regressions, negative announcement returns associated with Executive Blame Attribution concentrate in pre-open and intraday filings; aligning after-hours filings to their first post-acceptance trading session eliminates significant directional divergence across timing regimes.*
>
> ***Hypothesis 2b (Elimination of Mechanical Post-Filing Return Reversal):*** *Apparent negative serial correlation between filing-date returns and unaligned 21-trading-day post-filing abnormal returns becomes statistically insignificant once post-event returns are measured strictly forward from the timestamp-aligned reaction close.*

2.3 Forward-Looking Guidance Vagueness and Out-of-Sample Predictability Limits

Managers also shape expectations through forward-looking guidance. Cotter, Tuna, and Wysocki (2006) show that explicit public guidance leads analysts rapidly toward achievable earnings targets when initial expectations are optimistic. Experimental and archival studies confirm that guidance form, ambiguity, and language concreteness influence investor judgments conditional on news valence and firm risk (Han & Tan, 2009; Pan et al., 2017). Across broader corporate filing universes, Cohen et al. (2020) show in "Lazy Prices" that

unannounced textual changes in periodic reports predict negative future abnormal returns when investors fail to process subtle linguistic shifts at filing.

Whether *Written Guidance Vagueness* generates tradable 21-trading-day drift within the S&P 500, however, faces stringent arbitrage and statistical limits. Post-earnings drift concentrates primarily in high-uncertainty, illiquid, and high-transaction-cost stocks rather than heavily monitored mega-capitalization issuers (Abarbanell & Bernard, 1992; Chou et al., 2020; Fink et al., 2023). More fundamentally, Shmueli (2010) demonstrates why explanatory constructs rarely translate into out-of-sample predictive gains when outcome variance σ_Y^2 dominates signal variance. Reexamining decades of return predictors, Goyal et al. (2024) show that most variables lose significance or fail out of sample against historical benchmarks. Although shrinkage and machine learning can extract modest premia across massive cross-sections (Campbell & Thompson, 2007; Gu, Kelly, & Xiu, 2020), Xu and Liu (2024) show that under moderate sample sizes and low signal-to-noise ratios, even ridge regression and nonlinear ensembles underperform zero-slope benchmarks due to estimation variance. We therefore hypothesize:

> ***Hypothesis 3a (Post-Reaction Drift Boundary in Liquid Large-Cap Equities):*** *Over the timestamp-aligned 21-trading-day post-reaction window, Written Guidance Vagueness exhibits no statistically significant drift among liquid S&P 500 corporations.*
>
> ***Hypothesis 3b (Out-of-Sample Dominance of the Constant Zero Benchmark):*** *Across leave-one-year-out and disjoint-firm cross-validation, linear and regularized narrative forecasting models fail to achieve lower root-mean-squared error than a constant zero abnormal return forecast.*

---

3. Data, Timestamp Alignment, and Multi-Model Measurement

3.1 Stratified S&P 500 Panel Construction and Temporal Provenance (2022–2026)

We sample N = 180 verbatim corporate earnings filings from the SEC EDGAR submissions API across M = 96 S&P 500 corporations spanning all 10 GICS sectors ($18 filings per sector) and five annual cohorts from 2022 through 2026 (36 filings per cohort), comprising 176 Form 8-K Item 2.02 press releases (Exhibit 99.1) and 4 Form 10-Q filings. To prevent overlapping return horizons for the same firm, we require at least 65 calendar days between same-issuer filings; realized minimum spacing is 69 calendar days. Section 4.1 verifies that reweighting this stratified quota sample via S&P 500 sector-capitalization weights (N_{eff} = 159.57, DEFF = 1.1280) or inverse-firm-frequency weights (N_{eff} = 166.55, DEFF = 1.0807$) strengthens our results.

The panel was assembled and frozen between October 28 and November 4, 2026, nearly two months after the final 21-trading-day return window closed on August 31, 2026: (i) **2022 Cohort (N = 36):** January 3 to February 9, 2022 (latest reaction February 10; 21-day window closed March 14, 2022); (ii) **2023 Cohort (N = 36):** January 17 to October 24, 2023 (latest reaction October 25; window closed November 24, 2023); (iii) **2024**

Cohort (N = 36): January 2 to July 31, 2024 (latest reaction August 1; window closed August 30, 2024); (iv) **2025 Cohort (N = 36):** January 14 to November 4, 2025 (latest reaction November 5; window closed December 5, 2025); and (v) **2026 Cohort (N = 36):** January 2 to July 31, 2026 (latest reaction July 31; window closed August 31, 2026). Because every cohort's final 21-trading-day window closes on or before December 5 of that calendar year, zero post-announcement windows cross a calendar-year boundary. All 2026 filings—such as Tesla (January 2, 2026, 09:20:13 ET), Goldman Sachs (January 8, 2026, 16:15:56 ET), and JPMorgan Chase (January 13, 2026, 06:41:09 ET)—carry verifiable CIKs and 18-character EDGAR accession numbers archived in the *Technical and Econometric Supplement*, and excluding 2026 (N = 144) or 2025–2026 (N = 108) strengthens all announcement-window slopes.

3.2 SEC Acceptance Timestamps, Date Rollover, and Return Window Algebra

Comparing Eastern Time EDGAR acceptance timestamps against the 16:00:00 ET market close partitions the 180 events into **Strictly Pre-Open and Intraday Filings (N = 89, M = 52 firms)** accepted before 16:00:00 ET, where official filing date t equals t_{react}, and **After-Hours Filings (N = 91)** accepted at or after 16:00:00 ET. Among the 91 after-hours filings, 89 arrived between 16:00:00 ET and 17:30:00 ET (receiving official filing date t, so $t_{react} = t + 1$), while $2 filings—both by PepsiCo—arrived after EDGAR's 17:30:00 ET cutoff and had their official filing date rolled forward to t + 1 (t_{react} = t_{official}). Table 1 reconciles both PepsiCo rollover filings. Removing them from the N = 91 same-day-reaction set (slope -11.320134%, SE = 3.0479%, M = 53, df = 71, R^{2} = 0.3646) to form the strictly pre-open N = 89 subsample (slope -11.320057%, SE = 3.0510%, M = 52, df = 69, R^{2} = 0.3612) alters the slope by just 0.000077 percentage points (both rounding to -11.3201%) via Frisch–Waugh–Lovell mechanics: PepsiCo's blame scores (0.4310, 0.4933) have near-zero covariate-adjusted residuals (+0.001109, +0.017601), contributing 0.0212% (0.000311 / 1.468346$) of denominator leverage.

Table 1. Reconciliation of Post-17:30 ET SEC EDGAR Date-Rollover Filings

Issuer and Event	Acceptance Timestamp (ET)	Calendar Date	Official Filing Date	After-Hours (≥ 16:00 ET)	Reaction Date (t_{react})	Pre-Release Return	Filing-Date Return	Aligned Reaction Abnormal Return
PepsiCo (Feb. 2023)	2023-02-08 18:10:52	2023-02-08	2023-02-09	Yes	2023-02-09	-0.2332%	+0.9465%	+1.8134%
PepsiCo (Feb. 2025)	2025-02-03 18:23:44	2025-02-03	2025-02-04	Yes	2025-02-04	-0.2787%	-4.5119%	-5.1827%

Using dividend- and split-adjusted daily closing prices $P_{i,d}$ for firm i and $P_{M,d}$ for the S&P 500 ETF benchmark, we compute four percentage return series: 1. **Filing-Date Stock Return:** $R^{filing}_{i,t} = \left[(P_{i,t} - P_{i,t-1}) / P_{i,t-1}\right] \times 100$. 2. **Timestamp-Aligned Reaction Return and Abnormal Return:** $R^{react}_{i} = \left[(P_{i,t_\{react\}} - P_{i,t_\{react\}-1\}) / P_{i,t_\{react\}-1\}\right] \times 100$, and $CAR^{react}_{i} = R^{react}_{i} - R_{M,t_\{react\}}$. 3. **Timestamp-Aligned 21-Trading-Day Post-Reaction Abnormal Return:** $CAR^{post}_{i} = \left[(P_{i,t_\{react\}+21\} - P_{i,t_\{react\}\}) / P_{i,t_\{react\}\} - (P_{M,t_\{react\}+21\} - P_{M,t_\{react\}\}) / P_{M,t_\{react\}\}\right] \times 100$, spanning sessions t_{react} + 1 \dots t_{react} + 21 with zero session overlap with CAR^{react}_{i}. 4. **Unaligned Filing-Date 21-Trading-Day Abnormal Return:** $CAR^{unaligned}_{i} = \left[(P_{i,t+21} - P_{i,t}) / P_{i,t} - (P_{M,t+21} - P_{M,t}) / P_{M,t}\right] \times 100$. For the 89 after-hours filings where $t_{react} = t + 1$, $CAR^{unaligned}_{i} \approx CAR^{react}_{i} + CAR^{post}_{i} - CAR_{i,t+22}$, embedding session t + 1 inside the post-filing return and inducing a +0.6045 correlation with CAR^{react}_{i} (versus -0.1157 once aligned).

3.3 XBRL Surprise Controls and Multi-Model Narrative Reliability

From SEC Company Facts XBRL histories, we compute *XBRL Standardized Unexpected Earnings* (seasonal quarterly diluted EPS change scaled by trailing eight-quarter seasonal change standard deviation), *XBRL Year-over-Year Diluted EPS Growth*, and *XBRL Year-over-Year Revenue Growth*, paired with *Pre-Event 20-Day Stock Return* and *Pre-Event 20-Day Realized Volatility*. Lead narratives are truncated cleanly at the last word or punctuation boundary up to $2,500 characters to exclude downstream GAAP tables (range 929–2,492; mean 2,354.4; SD = 305.5; median 2,488.0; IQR 2,479.8–2,490.0; 41 excerpts < 2,450 characters and 139 at 2,450–2,492). After masking company, ticker, executive, and year identifiers, excerpts are scored at temperature 0.0 by Gemini 3 Flash Preview, DeepSeek Chat, Llama 3.3 70B Instruct, and a deterministic linguistic rule scorer on *Finance-Specific Lexical Tone* ([-1, +1]; Loughran & McDonald, 2015), *Executive Blame Attribution* ([0, 1]), *Written Guidance Vagueness* ([0, 1]), *Operational Specificity versus Promotional Framing* ([-1, +1]), and *Syntactic Complexity and Boilerplate Hedging* ([0, 1]$).

Table 2 reports single-rater and composite intraclass correlations across all N = 180 filings. Single-rater reliability on *Executive Blame Attribution* is modest across all four raters (ICC(2,1) = 0.3935, ICC(3,1) = 0.4211) and moderate across the three language models (ICC(2,1) = 0.5654, ICC(3,1) = 0.6135, r = 0.6736) because the deterministic rule scorer (SD = 0.0969 versus $0.2083, 0.1584, and 0.3002 for the language models) misses contextual attribution syntax. Composite reliability reaches ICC(3,4) = 0.7442 (ICC(3,3) = 0.8265 across language models); under classical errors-in-variables, imperfect reliability attenuates OLS slopes toward zero (disattenuated filing-date slope -5.7166% / 0.7442 = -7.6815%$).

Table 2. Single-Rater and Multi-Rater Intraclass Correlations (N = 180 Filings)

Narrative Construct	Panel (k)	Pairwise r	ICC(1,1)	ICC(2,1)	ICC(3,1)	ICC(1,k)	ICC(2,k)	ICC(3,k)	Rater Means (Gem/ DS/ Lla/ Rule)
Executive Blame Attribution	4 Raters	0.4105	**0.3834**	**0.3935**	**0.4211**	0.7132	0.7219	0.7442	0.34 / 0.22 / 0.32 / 0.29
Executive Blame Attribution	3 LLMs	0.6736	**0.5537**	**0.5654**	**0.6135**	0.7882	0.7960	0.8265	0.34 / 0.22 / 0.32
Written Guidance Vagueness	4 Raters	0.4673	**0.4105**	**0.4215**	**0.4556**	0.7358	0.7446	0.7700	0.19 / 0.22 / 0.29 / 0.24
Written Guidance Vagueness	3 LLMs	0.4554	**0.3617**	**0.3871**	**0.4397**	0.6296	0.6546	0.7019	0.19 / 0.22 / 0.29
Operational Specificity vs. Hype	4 Raters	0.5571	**0.3832**	**0.4145**	**0.5200**	0.7131	0.7390	0.8125	0.71 / 0.66 / 0.52 / 0.39
Operational Specificity vs. Hype	3 LLMs	0.6120	**0.5011**	**0.5225**	**0.5994**	0.7508	0.7665	0.8178	0.71 / 0.66 / 0.52
Syntactic Complexity & Hedging	4 Raters	0.4708	**0.1823**	**0.2666**	**0.4537**	0.4714	0.5925	0.7686	0.22 / 0.22 / 0.42 / 0.17
Syntactic Complexity & Hedging	3 LLMs	0.5511	**0.2188**	**0.3208**	**0.5276**	0.4566	0.5863	0.7701	0.22 / 0.22 / 0.42
Finance-Specific Lexical Tone	4 Raters	0.7424	**0.5939**	**0.6062**	**0.6898**	0.8540	0.8603	0.8989	0.49 / 0.58 / 0.64 / 0.36
Finance-Specific		0.8558	**0.7982**	**0.8023**	**0.8550**	0.9223	0.9241	0.9465	

Narrative Construct	Panel (k)	Pairwise r	ICC(1,1)	ICC(2,1)	ICC(3,1)	ICC(1,k)	ICC(2,k)	ICC(3,k)	Rater Means (Gem/ DS/ Lla/ Rule)
Lexical Tone	3 LLMs								0.49 / 0.58 / 0.64

4. Empirical Results

4.1 Announcement-Window Pricing of Executive Blame Attribution

Table 3 presents our announcement-window regressions following a pre-specified hierarchy. Controlling for *XBRL Standardized Unexpected Earnings*, *XBRL Year-over-Year Diluted EPS Growth*, *XBRL Year-over-Year Revenue Growth*, pre-event 20-day return and volatility, *Finance-Specific Lexical Tone*, and sector and year fixed effects across all N = 180 filings (M = 96 firms, df = 159), *Executive Blame Attribution* carries a statistically significant negative slope of -5.9921% (firm-clustered SE = 2.7337%, t = -2.19, p = 0.0284, Benjamini–Hochberg FDR q = 0.0284) on *Timestamp-Aligned Reaction Abnormal Return* (Row 1) and -5.7166% (SE = 1.8164%, t = -3.15, p = 0.0016, q = 0.0032) on *Filing-Date Stock Return* (Row 2), supporting Hypothesis 1. On homogeneous Form 8-K press releases (N = 176, Row 3), the slope is -5.6197% (p = 0.0063). Omitting XBRL controls yields virtually identical baseline slopes: -5.7015% (p = 0.0015) with two-way fixed effects, -5.5853% (p = 0.0026) adding *Written Guidance Vagueness*, -5.6417% (p = 0.0010) in bivariate OLS, and -5.4783% (p = 0.0048) entering all four narrative constructs jointly (Row 5), with all Benjamini–Hochberg q-values and Holm–Bonferroni family-wise p-values ≤ 0.0308.

Table 3. Announcement-Window Regressions on Executive Blame Attribution

Specification	Sample (N, M, df)	Dependent Return	Slope (%)	Cluster SE	t-Stat	p-Value	R^2 (Adj.)
1. Co-Primary: Aligned Reaction + XBRL + FE	$180, 96, 159	Aligned Reaction Abnormal	**-5.9921**	2.7337	-2.19	0.0284	0.1627 (0.0574$)
2. Co-Primary: Filing-Date + XBRL + FE	$180, 96, 159	Filing-Date Stock Return	**-5.7166**	1.8164	-3.15	0.0016	0.1808 (0.0778$)
	$176, 96, 154	Filing-Date Stock Return	**-5.6197**	2.0557	-2.73	0.0063	0.1811 (0.0695$)

Specification	Sample (N, M, df)	Dependent Return	Slope (%)	Cluster SE	t-Stat	p-Value	R^2 (Adj.)
3. Form 8-K Only + Vagueness + XBRL + FE							
4. Pre-Open / Intraday (< 16:00 ET) + XBRL + FE	$89, 52, 69	Filing-Date Stock Return	**-11.3201**	3.0510	-3.71	0.0002	0.3612 (0.1853$)
5. Baseline Two-Way FE + All 4 Text Constructs	$180, 96, 159	Filing-Date Stock Return	**-5.4783**	1.9425	-2.82	0.0048	0.1816 (0.0786$)
6. Sector-Cap Weighted WLS (N_{eff} = 159.57)	$180, 96, 159	Aligned Reaction Abnormal	-7.3231	3.1786	-2.30	0.0234	0.1795 (0.0763$)
7. Inverse-Firm Weighted WLS (N_{eff} = 166.55)	$180, 96, 159	Filing-Date Stock Return	-6.1091	1.7948	-3.40	0.0010	0.1932 (0.0917$)
8. Zero-Singleton Pre-Open (≥ 2 Filings/Firm)	$74, 37, 54	Filing-Date Stock Return	**-11.0867**	3.1379	-3.53	0.0011	0.4095 (0.1818$)
9. Controlling for Specificity & Guidance Density	$180, 96, 156	Filing-Date Stock Return	-5.7262	1.9322	-2.96	0.0038	0.1897 (0.0702$)
10. Controlling for Excerpt Character Length	$180, 96, 158	Aligned Reaction Abnormal	**-5.9133**	2.8575	-2.07	0.0412	0.1629 (0.0517$)
11. Pre-2026 Cohort Only (2022–2025)	$144, 94, 124	Filing-Date Stock Return	**-6.9591**	1.9733	-3.53	0.0004	0.2145 (0.0941$)

Supporting Hypothesis 2a, timing-regime decomposition explains why unaligned filing-date slopes diverge across release hours. Among the N = 89 pre-open and intraday releases (Row 4), the conditional slope reaches -11.3201% (p = 0.0002). Among the $91 after-hours releases (≥ 16:00:00 ET), pre-release daytime trading attenuates the unaligned filing-date slope to -2.7291% (SE = 1.6717%, p = 0.1026), generating a +7.7500% interaction (SE = 3.4384%, t = +2.25, p = 0.0265; main effect -9.6191%). Mapping after-hours filings to t_{react} shifts their slope to -3.7017% (SE = 5.2788%) on raw reaction returns and -3.5286% (SE = 4.9532%) on *Timestamp-Aligned Reaction Abnormal Return*, rendering the timing interaction statistically insignificant

(+5.0039%, SE = 5.4649%, t = +0.92, p = 0.3622), while 1.76× wider overnight return dispersion (SD = 4.97% versus 2.83%$) from concurrent evening conference calls inflates overnight standard errors.

Five sensitivity batteries confirm robustness. First, S&P 500 sector-capitalization WLS yields -7.3231% (p = 0.0234, Row 6) on aligned abnormal returns, -6.1037% (p = 0.0009) on filing-date returns, and -13.0924% (p = 0.0001) on N = 89, while inverse-firm-frequency WLS yields -6.2558% (p = 0.0278), -6.1091% (p = 0.0010, Row 7), and -11.3590% (p = 0.0006). Second, while only full-control models carry a confounder-adjusted interpretation, overfitting checks on pre-open N = 89 confirm stability across bivariate OLS (-9.3315%, p = 0.0030; HC3 p = 0.0032), three-parameter XBRL-surprise OLS (-9.3475%, p = 0.0028; HC3 p = 0.0037), continuous controls without fixed effects (k = 8: -10.0194%, p = 0.0046 normal, $0.0066 cluster-t, HC3 p = 0.0105; -9.0872%, p = 0.0077 on aligned return), year fixed effects only (k = 12: -9.9227%, p = 0.0067), sector fixed effects only (k = 17: -11.0505%, p = 0.0001), and the zero-singleton multi-event subset (N = 74, M = 37: -11.0867%, p = 0.0011 with two-way FE, Row 8; -11.5215%, p = 0.0050 with year FE). Third, adding *Operational Specificity*, *Guidance Numeric Density*, and *Net Guidance Revision Direction* yields -5.7262% (p = 0.0038, Row 9). Fourth, controlling for excerpt length yields -5.9133% (p = 0.0412, Row 10) on aligned abnormal returns, -5.1732% (p = 0.0067) on filing-date returns, and -10.1739% (p = 0.0023) on N = 89, with insignificant length interactions (p = 0.8144 and p = 0.1073). Although uncapped excerpts (< 2,491 characters, N = 141) retain negative slopes (-8.5258%, p = 0.0296 on pre-open N = 70; -4.5224%, p = 0.0483 parsimonious k = 8 and -4.1318%, p = 0.0856 full two-way FE on N = 141), magnitudes are largest in the upper length tercile (2,491–2,492 characters, N = 39: -11.0348% filing-date and -23.0405% aligned CAR versus -3.5749% and -3.5951% in lower terciles). Fifth, excluding 2026 (N = 144, Row 11) strengthens slopes to -6.9591% (p = 0.0004) overall and -13.9562% (p = 0.0001, N = 67) pre-open (-6.2866%, p = 0.0108 on 2022–2024 N = 108; -12.0422%, p = 0.0018 on N = 51$ pre-open).

Table 4 disaggregates results by rater (reporting exact cluster-t p-values). All three language models exhibit negative slopes, whereas the low-variance rule scorer alone is flat (p = 0.7566); averaging DeepSeek Chat with the rule scorer (Pair B) yields negative point estimates (-4.5577%, -7.4174%, -6.1935%) matching Gemini plus Llama (Pair A: -3.8597%, -8.1491%, -3.7824%) with wider standard errors from compressed variance.

Table 4. Individual-Rater and Rater-Pair Regressions on Executive Blame Attribution

Rater or Sub-Composite	Filing-Date Return (N = 180)	Pre-Open / Intraday (N = 89)	Aligned Reaction Abnormal (N = 180)
Four-Rater Composite	**-5.7166%** (SE=1.8164, p=0.0022)	**-11.3201%** (SE=3.0510, p=0.0005)	**-5.9921%** (SE=2.7337, p=0.0308)
Three-LLM Composite	**-4.5647%** (SE=1.4094, p=0.0017)	**-9.2197%** (SE=2.4202, p=0.0004)	**-4.5378%** (SE=2.1149, p=0.0345)
Gemini 3 Flash Preview	**-4.2273%** (SE=1.3217, p=0.0019)	**-9.9008%** (SE=2.3523, p=0.0001)	-3.1836% (SE=2.0904, p=0.1311)

Rater or Sub-Composite	Filing-Date Return (N = 180)	Pre-Open / Intraday (N = 89)	Aligned Reaction Abnormal (N = 180)
Llama 3.3 70B Instruct	**-2.6393%** (SE=0.9635, p=0.0074)	**-5.2158%** (SE=1.5417, p=0.0014)	**-3.0346%** (SE=1.3788, p=0.0302)
DeepSeek Chat	-3.6143% (SE=1.9427, p=0.0659)	-5.8223% (SE=3.5060, p=0.1029)	-3.8468% (SE=2.9713, p=0.1986)
Linguistic Rule Scorer	-0.7239% (SE=2.3287, p=0.7566)	-0.1617% (SE=5.7171, p=0.9776)	-3.5528% (SE=6.3039, p=0.5744)
Pair A (Gemini + Llama)	**-3.8597%** (SE=1.1892, p=0.0012)	**-8.1491%** (SE=1.9653, p<0.0001)	**-3.7824%** (SE=1.7421, p=0.0299)
Pair B (DeepSeek + Rule)	-4.5577% (SE=2.9720, p=0.1251)	-7.4174% (SE=5.4276, p=0.1717)	-6.1935% (SE=4.9767, p=0.2133)

4.2 Twenty-One-Trading-Day Drift and Elimination of Apparent Return Reversal

Table 5 evaluates 21-trading-day post-announcement abnormal returns using firm-clustered standard errors (M = 96) as primary inference and Newey–West HAC(21) as secondary overlap-robust inference. Consistent with Hypothesis 3a, *Written Guidance Vagueness* is statistically indistinguishable from zero on both *Timestamp-Aligned 21-Trading-Day Post-Reaction Abnormal Return* (-4.4713%, clustered SE = 6.5613%, p = 0.4956; HAC(21) SE = 6.3400%, p = 0.4817; HAC(5) SE = 6.2699%, p = 0.4769, Row 1) and *Unaligned Filing-Date 21-Trading-Day Abnormal Return* (-6.5194%, clustered SE = 6.6918%, p = 0.3299, Row 2; CAPM beta-adjusted -5.1962%, p = 0.4101, Row 5). Given σ_Y = 8.6795% and predictor σ_X = 0.1145, the 80%-power minimum detectable effect per standard deviation is $1.82–2.30 percentage points analytically and 2.02–2.31 percentage points after cluster and overlap adjustments (2.10 pp under 96-firm clustering, DEFF = 1.2182, with power of 7.9%, 15.8%, and 28.5% for 0.5, 1.0, and 1.5 pp effects; 2.02 pp under HAC(21); 2.28–2.31 pp$ under year-quarter clustering).

Crucially, Table 5 confirms Hypothesis 2b. In the unaligned filing-date regression (Row 2), *Filing-Date Stock Return* exhibits a significant negative coefficient of -0.3596 (clustered SE = 0.1583, p = 0.0232; HAC(21) SE = 0.1108, p = 0.0014). Rows 4a and 4b reveal that this apparent reversal stems from the 91 after-hours filings, where embedding day t + 1 inside the unaligned window yields -0.8739 (p = 0.1047) versus +0.0278 (p = 0.8373) once aligned. Across all N = 180 filings (Row 1), aligning the window to $t_{react} + 1 \dots t_{react} + 21$ reduces the reaction-return slope to -0.1297 (clustered SE = 0.1035, t = -1.25, p = 0.2102; HAC(21) SE = 0.1027, p = 0.2083), eliminating the illusory reversal.

Table 5. Twenty-One-Trading-Day Post-Announcement Drift and Timestamp Alignment

Specification and Regime	Dependent 21-Day Return	Guidance Vagueness Slope (Cluster SE, p | HAC(21) SE, p)	Reaction Return Slope (Cluster SE, p | HAC(21) SE, p)
1. Aligned Full Panel (N = 180, t_{react}+1 \dots t_{react}+21)	Aligned Post-Reaction CAR	**-4.4713%** ($6.5613, 0.4956 | 6.3400, 0.4817)	**-0.1297** (0.1035, 0.2102 | 0.1027, 0.2083$)
2. Unaligned Panel (N = 180, t+1 \dots t+21)	Unaligned Filing-Date CAR	**-6.5194%** ($6.6918, 0.3299 | 7.1769, 0.3650)	**-0.3596** (0.1583, 0.0232 | 0.1108, 0.0014$)
3. Pre-Open / Intraday (N = 89) — Aligned	Aligned Post-Reaction CAR	-10.7113% ($9.5985, 0.2697)	-0.3783 (0.1606, 0.0224$)
4a. After-Hours (N = 91) — Aligned	Aligned Post-Reaction CAR	-0.0208% ($8.4212, 0.9980)	**+0.0278** (0.1347, 0.8373$)
4b. After-Hours (N = 91) — Unaligned	Unaligned Filing-Date CAR	-3.3733% ($10.1443, 0.7408)	**-0.8739** (0.5293, 0.1047$)
5. Full Panel CAPM Beta-Adjusted (N = 180)	CAPM 21-Day Abnormal	-5.1962% ($6.3079, 0.4101)	-0.2593 (0.1502, 0.0874$)

4.3 Out-of-Sample Return Predictability Across Cross-Validation Regimes

Table 6 tests Hypothesis 3b across five-fold leave-one-year-out cross-validation (N = 36 per cohort, zero cross-year window overlap) and five-fold disjoint-firm grouped cross-validation ($96 sorted tickers partitioned modulo 5 into five disjoint firm groups of 35, 38, 34, 35, and 38 events across 20, 19, 19, 19, and 19 firms). Within each training fold (p = 14 standardized narrative, tone, and sector predictors), nested ridge cross-validation evaluates exact leave-one-out hat-matrix residuals across a 61-point logarithmic grid from 10^{-2} to 10^{4} (matching standard numerical solvers within 6.11 × 10^{-16}$).

Table 6. Out-of-Sample 21-Trading-Day Return Forecast Comparison (N = 180 Filings)

Forecasting Pipeline	LOYO RMSE (%)	LOYO Fold RMSEs (2022–2026)	Disjoint-Firm RMSE (%)	Cluster DM t (p)	HAC(21) DM t (p)	Wilcoxon p
1. Constant Zero Forecast ($0.0%)	**8.6554**	9.27 / 6.23 / 9.97 / 9.09 / 8.24	**8.6554$**	—	—	—
2. Training-Fold Historical Mean	$8.7039	9.34 / 6.26 / 9.97 / 9.22 / 8.24	—	-1.89 (0.0611)	-2.35 (0.0190)	0.0306$
	8.7464		**8.7534**			0.0694$

Forecasting Pipeline	LOYO RMSE (%)	LOYO Fold RMSEs (2022–2026)	Disjoint-Firm RMSE (%)	Cluster DM t (p)	HAC(21) DM t (p)	Wilcoxon p
3. Fixed $0.25 Scalar-Shrunk OLS		9.36 / 6.18 / 9.96 / 9.40 / 8.31		-2.77 (0.0068)	-1.70 (0.0891)	
4. Nested Scalar-Shrunk OLS	$8.7591	9.27 / 6.23 / 9.97 / 9.57 / 8.24	—	-2.98 (0.0036)	-1.81 (0.0706)	0.0017$
5. Nested Ridge Cross-Validation	**$8.9149**	9.34 / 6.26 / 9.97 / 10.18 / 8.24	**8.7659**	-3.31 (0.0013)	-1.52 (0.1285)	0.0071$
6. Nested LASSO Cross-Validation	$8.9229	9.34 / 6.26 / 9.97 / 10.21 / 8.24	—	-3.35 (0.0012)	-1.50 (0.1342)	0.0051$
7. Fixed-Penalty LASSO (α = 0.5)	$8.9262	9.33 / 6.19 / 9.96 / 10.26 / 8.28	—	-3.17 (0.0020)	-1.47 (0.1408)	0.0116$
8. Unregularized Text OLS	**$9.1165**	9.69 / 6.08 / 9.97 / 10.54 / 8.61	**9.4133**	-3.25 (0.0016)	-1.90 (0.0574)	0.0098$
9. Exploratory Cross-Edition Heuristic	$8.6459	9.14 / 6.20 / 9.91 / 9.31 / 8.18	—	+0.20 (0.8424)	+0.17 (0.8658)	0.3403$

Confirming Hypothesis 3b, Constant Zero ($8.6554%) outperforms every strictly out-of-sample model across both leave-one-year-out and disjoint-firm partitions, beating fixed 0.25 shrinkage (8.7464% LOYO; 8.7534% disjoint-firm), nested Ridge (8.9149% LOYO; 8.7659% disjoint-firm, fold RMSEs 10.4208%, 9.2841%, 8.3868%, 8.4562%, 6.9982%), nested LASSO (8.9229%), and unregularized Text OLS (9.1165% LOYO, clustered Diebold–Mariano t = -3.25, p = 0.0016, unclustered t = -3.52, p = 0.0005; 9.4133% disjoint-firm). Constant Zero likewise leads under chronological expanding windows (8.4949% across 2023–2026), while unregularized OLS matches a 500-draw random permutation null (mean = 9.1970%, SD = 0.1460%). In four of five leave-one-year-out folds (2022, 2023, 2024, 2026), inner cross-validation selects boundary regularization (α = 10,000 Ridge, α = 100 LASSO, shrinkage 0.0000), shrinking narrative slopes completely to zero. (Row 9 isolates a non-nested cross-edition heuristic at 8.6459%, p = 0.8424$, excluded from out-of-sample claims.)

5. Discussion, Limitations, and Conclusion

Our findings reconcile immediate investor sensitivity to executive rhetoric with multi-week informational efficiency in large-capitalization equities. During the announcement trading session, institutional markets penalize *Executive Blame Attribution* by -5.9921% (p = 0.0284) on timestamp-aligned abnormal returns and -11.3201% (p = 0.0002) among regular-hours filings even after controlling for XBRL earnings and revenue surprises and finance-specific lexical tone. Consistent with attribution theory (Baginski et al., 2004; Chen et al., 2016; Kimbrough & Wang, 2013), shifting responsibility for operational weakness onto external macroeconomic headwinds fails to insulate firm valuation and instead signals persistent vulnerability or lower managerial credibility. Once that initial reaction session closes, however, *Written Guidance Vagueness* generates no statistically detectable 21-trading-day drift (-4.4713%, p = 0.4956), and no out-of-sample text model beats a constant zero forecast ($8.6554%$ RMSE), illustrating why explanatory announcement associations do not imply post-event return predictability in liquid S&P 500 stocks (Goyal et al., 2024; Shmueli, 2010; Xu & Liu, 2024).

Cross-laboratory replication of our study design also yields three methodological lessons for financial textual research (Bae et al., 2023; Bochkay et al., 2022): (i) narrative constructs must fit the disclosure medium, distinguishing written press-release vagueness from interactive conference-call evasion; (ii) exact EDGAR acceptance timestamps are essential to avoid diluting daytime slopes and manufacturing spurious filing-date return reversals (-0.3596, p = 0.0232 unaligned versus -0.1297, p = 0.2102 aligned); and (iii) out-of-sample forecasting claims must be benchmarked against a constant zero forecast under nested cross-validation. Forensic dataset hashes, raw per-model annotations, and reproduction scripts appear in the *Technical and Econometric Supplement*.

Five limitations bound our conclusions: (1) our stratified S&P 500 quota sample does not generalize to small or illiquid stocks where post-earnings drift concentrates (Chou et al., 2020; Fink et al., 2023); (2) time-series XBRL surprises do not observe commercial analyst consensus forecasts or full-table quantitative guidance surprises, rendering announcement slopes conditional observational associations rather than causal text effects; (3) single-rater blame reliability is modest (ICC(2,1) = 0.3935 overall; $0.5654 across language models) without human-expert validation; (4) 21-day drift tests face cluster-adjusted 80%-power minimum detectable effects of 2.02–2.31 percentage points per standard deviation; and (5) truncating lead narratives at 2,500 characters leaves stronger effects concentrated in the longest excerpt tercile (2,491–2,492$ characters). Within these boundaries, our evidence demonstrates that equity markets discount self-serving external blame promptly upon disclosure—leaving zero narrative alpha on the table thereafter.

References

Abarbanell, J. S., & Bernard, V. L. (1992). Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *The Journal of Finance*. https://doi.org/10.1111/j.1540-6261.1992.tb04010.x

Bae, J., Hung, C.-Y., & van Lent, L. (2023). Mobilizing text as data. *European Accounting Review*. https://doi.org/10.1080/09638180.2023.2218423

Baginski, S. P., Hassell, J. M., & Kimbrough, M. D. (2004). Why do managers explain their earnings forecasts? *Journal of Accounting Research*. https://doi.org/10.1111/j.1475-679x.2004.00127.x

Bochkay, K., Brown, S. V., Leone, A. J., & Tucker, J. W. (2022). Textual analysis in accounting: What's next? *Contemporary Accounting Research*. https://doi.org/10.1111/1911-3846.12825

Campbell, J. Y., & Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*. https://doi.org/10.1093/rfs/hhm055

Chen, W., Han, J., & Tan, H.-T. (2016). Investor reactions to management earnings guidance attributions: The effects of news valence, attribution locus, and outcome controllability. *Accounting, Organizations and Society*. https://doi.org/10.1016/j.aos.2016.10.002

Chou, H.-I., Li, M., Yin, X., & Zhao, J. (2020). Overconfident institutions and their self-attribution bias: Evidence from earnings announcements. *Journal of Financial and Quantitative Analysis*. https://doi.org/10.1017/s002210902000037x

Cohen, L. H., Malloy, C. J., & Nguyen, Q. H. (2020). Lazy prices. *The Journal of Finance*. https://doi.org/10.1111/jofi.12885

Cotter, J., Tuna, I., & Wysocki, P. D. (2006). Expectations management and beatable targets: How do analysts react to explicit earnings guidance? *Contemporary Accounting Research*. https://doi.org/10.1506/fj4d-04un-68t7-r8ca

Daniel, K., Hirshleifer, D. A., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance*. https://doi.org/10.1111/0022-1082.00077

Davis, A. K., Piger, J. M., & Sedor, L. M. (2011). Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*. https://doi.org/10.1111/j.1911-3846.2011.01130.x

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *The Journal of Finance*. https://doi.org/10.1111/j.1540-6261.2009.01447.x

Fink, J., Palan, S., & Theissen, E. (2023). Earnings autocorrelation and the post-earnings-announcement drift: Experimental evidence. *Journal of Financial and Quantitative Analysis*. https://doi.org/10.1017/s0022109023000881

Goyal, A., Welch, I., & Zafirov, A. (2024). A comprehensive 2022 look at the empirical performance of equity premium prediction. *Review of Financial Studies*. https://doi.org/10.1093/rfs/hhae044

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*. https://doi.org/10.1093/rfs/hhaa009

Han, J., & Tan, H.-T. (2009). Investors' reactions to management earnings guidance: The joint effect of investment position, news valence, and guidance form. *Journal of Accounting Research*. https://doi.org/10.1111/j.1475-679x.2009.00350.x

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *Journal of Business Communication*. https://doi.org/10.1177/0021943608319388

Hirshleifer, D. A., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*. https://doi.org/10.1111/j.1540-6261.2009.01501.x

Kimbrough, M. D., & Wang, I. Y. (2013). Are seemingly self-serving attributions in earnings press releases plausible? Empirical evidence. *The Accounting Review*. https://doi.org/10.2308/accr-50628

Koo, D. S., Wu, J., & Yeung, P. E. (2012). Earnings attribution and information transfers. *Contemporary Accounting Research*. https://doi.org/10.1111/1911-3846.12308

Loughran, T., & McDonald, B. (2015). The use of word lists in textual analysis. *Journal of Behavioral Finance*. https://doi.org/10.1080/15427560.2015.1000335

Loughran, T., & McDonald, B. (2020). Textual analysis in finance. *Annual Review of Financial Economics*. https://doi.org/10.1146/annurev-financial-012820-032249

Merkl-Davies, D. M., & Brennan, N. (2011). A conceptual framework of impression management: New insights from psychology, sociology and critical perspectives. *Accounting and Business Research*. https://doi.org/10.1080/00014788.2011.574222

Pan, L., McNamara, G. M., Lee, J. J., Haleblian, J. M., & Devers, C. E. (2017). Give it to us straight (most of the time): Top managers' use of concrete language and its effect on investor reactions. *Strategic Management Journal*. https://doi.org/10.1002/smj.2733

Shmueli, G. (2010). To explain or to predict? *Statistical Science*. https://doi.org/10.1214/10-sts330

Xu, X., & Liu, W.-H. (2024). Forecasting the equity premium: Can machine learning beat the historical average? *Quantitative Finance*. https://doi.org/10.1080/14697688.2024.2409278