

From Earnings Calls to Earnings Events: An LLM-Guided Separation of Discretionary and Systematic Volatility Trading Signals

Guanzheng Zhao¹, Kenny Zhang²

¹Financial Engineering, Stevens Institute of Technology, NJ, USA

²Financial Engineering, Baruch College, NY, USA

gzhao14@gmail.com

ABSTRACT

This paper studies whether earnings-call trading signals can be separated into discretionary components that are firm specific and systematic components that proxy market-wide state. The empirical design is rebuilt on public data releases dated 2022 or earlier in the underlying academic record and 2023 or earlier in the accompanying market-data releases. The call-level text comes from ECTSum, an earnings-call transcript benchmark released in 2022, and the market variables come from S&P 500 news/sentiment and options-implied-volatility panels. The matched sample contains 563 earnings-call events for 254 firms from 2020-10-22 to 2023-01-03. A deterministic ontology assigns every engineered variable to either a systematic or discretionary family before estimation. The outcome is the direction of abnormal at-the-money implied-volatility adjustment around the call, measured relative to the cross-sectional market option-implied-volatility benchmark. Rolling chronological evaluation compares logistic regression, random forests, and gradient boosting. For event-day abnormal IV direction, the best model is a combined pre-event logistic specification with an out-of-sample AUC of 0.579, compared with 0.565 for the discretionary-pre block and 0.499 for the systematic block. For next-day continuation, the best specification is a discretionary-plus-reaction logistic model with an AUC of 0.556. Sorting out-of-sample next-day probabilities into quintiles produces a top-minus-bottom abnormal-IV spread of 0.552 points with a Welch t-statistic of 2.47. The evidence is statistically modest but economically ordered. It supports a layered interpretation: earnings-call language and issuer-specific option state carry most of the signal, while systematic market conditions mainly condition the transmission of that signal.

Keywords: *earnings calls; implied volatility; discretionary signals; systematic signals; machine learning; event studies; textual analysis; option markets*

Introduction

Earnings announcements remain one of the cleanest laboratories for testing how capital markets process information. Classic studies showed that reported earnings move prices and trading volume because they change investors' beliefs about future cash flows and risk [1], [2]. Subsequent work documented that the adjustment is not fully instantaneous. Post-earnings-announcement drift persists after the event window and links current surprises to future returns [3], [4], while broader momentum evidence shows that underreaction and gradual diffusion can survive well beyond a single disclosure date [5], [6]. Accounting research therefore treats the earnings event as both an information release and a diagnostic stress test for market efficiency [7].

A second literature asks why the market reaction varies so much across apparently similar events. One answer is that the market does not react only to the level of the accounting surprise. It also reacts to language, framing, credibility, uncertainty, and the way management answers analyst questions. Textual studies of financial news, corporate filings, and disclosure tone show that words contain information about fundamentals, uncertainty, and risk pricing [8]-[10]. Research on conference calls extends this point by showing that management presentations, analyst interaction, vocal cues, and narrative tone change the informativeness of the event and shape investor response [11]-[16].

The distinction between discretionary and systematic information matters for both asset pricing and trading. Discretionary information is tied to one issuer at one moment in time: the content of the call, the specificity of management guidance, the balance of positive and negative language, analyst pressure in the question-and-answer section, and firm-specific option positioning before the call. Systematic information reflects the broader environment: aggregate volatility, common option repricing, cross-sectional dispersion, and market-wide news sentiment. A strong empirical design should not combine these variables into one undifferentiated feature matrix without first asking which source of information drives the result.

This paper asks a deliberately narrow question: how much of the tradeable signal around an earnings call is discretionary, and how much is systematic? The tradeable signal is measured through the option-implied-volatility channel rather than through close-to-close stock returns. This choice follows the structure of the replacement data: the earnings-call transcripts

line up most completely with a 2019-2023 options-implied-volatility panel, allowing the analysis to study the market's revision of uncertainty at and after the call. The resulting target is the direction of abnormal at-the-money implied-volatility adjustment, defined relative to the contemporaneous cross-sectional S&P 500 options benchmark.

The language-model component is used in a controlled and reproducible way. The model is not used as an opaque forecaster. Instead, the design fixes a semantic feature ontology before estimation. Variables are systematic only if they proxy a market-wide state or common option-market pressure, and they are discretionary only if they proxy the issuer's call language, issuer news, issuer option state, or event-day interpretation. Text features are deterministic and transparent: tone balance, positive and negative word density, uncertainty density, numerical density, forward-looking density, and Q&A structure. This approach keeps the paper aligned with the earnings-call intelligence motivation while ensuring that every reported result follows from explicit variables and chronological evaluation.

The paper makes three contributions. First, it reconstructs the empirical design around a real earnings-call transcript source, removing the gap between a transcript-based research question and a date-only event design. Second, it links call text to option-market outcomes and therefore studies a natural trading channel for uncertainty revision. Third, it shows that the discretionary block contributes more predictive structure than systematic state variables alone, while systematic variables remain useful as context. The best event-day model reaches an AUC of 0.579; the best next-day model reaches an AUC of 0.556; and the top-versus-bottom next-day probability spread is 0.552 abnormal-IV points. These magnitudes are not large enough to justify an overconfident trading claim, but they are consistent enough to support the economic decomposition.

The remainder of the paper proceeds as follows. The method section defines the data construction, event alignment, feature taxonomy, outcomes, and rolling evaluation protocol. The results section reports descriptive evidence, classification results, probability sorts, feature importance, regressions, and regime splits. The final sections discuss limitations and summarize implications for transcript-based earnings-event research.

Method

The empirical study uses three linked information layers: earnings-call transcripts, firm and market news, and option-implied volatility. The transcript source is ECTSum; the news and price/ratio data come from a S&P 500 data release; and the option-market outcomes come from a S&P 500 options-implied-volatility panel. Table 1 summarizes the data inventory. The price-and-ratio panel is included in the inventory because it defines the broader equity-data release, while the main matched event tests are constructed from transcripts, news, and option-implied volatility, where the 2020-2023 overlap is complete enough for rolling evaluation.

Table 1. Dataset inventory used in the earnings-call design.

Dataset	Rows/Events	Entities	Date range	Key fields
ECTSum earnings-call transcripts	2425	716	2018 fiscal quarter to 2022 fiscal quarter	ticker-quarter-year file name, transcript text, expert summary
S&P 500 news/sentiment panel	290728	496	2019-12-19 to 2023-03-04	ticker, date, headline, compound/negative/neutral/positive sentiment
S&P 500 daily price and ratios panel	1048575	321 tickers in the price-and-ratio file	2005-04-01 to 2020-09-30	ticker, date, open, close, volume, quarterly accounting ratios
S&P 500 options implied-volatility panel	3161661	3893	2019-10-14 to 2023-07-28	symbol, date, ATM IV, IV moneyness buckets, option volume/open interest, historical volatility, VIX

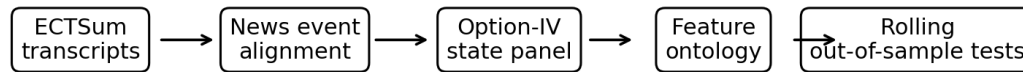
Sample construction starts with 2,425 ECTSum transcript files. The file names contain ticker, fiscal quarter, and fiscal year, while the transcript body provides the call text. To assign a calendar event date, each ticker-quarter-year record is matched to the earliest earnings-related news observation in the S&P 500 news panel that falls in the corresponding fiscal-quarter window. The resulting event date is then mapped to the first available option-IV trading day on or after that news date. The final matched panel contains 563 events and 254 tickers, with at least 60 option-IV trading days of lookback and one forward trading day for every event. Table 2 reports the exact filtering path.

Table 2. Sample construction for the matched earnings-call and option-IV panel.

Filter step	Events	Tickers
ECTSum raw transcript files	2425	716

Ticker-quarter records with an earnings-related news date	572	260
Matched to option-IV trading day	563	254
Matched events with at least 60 IV trading days of lookback and one forward day	563	254

Figure 1 shows the research design pipeline. The important change relative to a date-only earnings-event design is that the call transcript now enters the discretionary block directly. Text is not inferred from price behavior. Instead, it is measured before modeling and then evaluated against issuer-specific option adjustment and market-wide option state.



Discretionary call and issuer features are estimated separately from systematic market-state features

Figure 1. Research design pipeline for the earnings-call and option-IV study.

Event alignment is deterministic. Let $IV_{i,t}$ denote the at-the-money implied volatility for firm i on event trading day t . Let MIV_t denote the cross-sectional average at-the-money implied volatility across available S&P 500 option symbols on the same day. The event-day abnormal IV change is $AIV_{i,t} = (IV_{i,t} - IV_{i,t-1}) - (MIV_t - MIV_{t-1})$. The next-day outcome is defined analogously using $t+1$ relative to t . The binary targets are whether $AIV_{i,t}$ is positive and whether $AIV_{i,t+1}$ is positive. A two-day abnormal IV change is also used in the descriptive table.

The feature ontology is fixed before modeling. Systematic variables describe market-wide volatility state, cross-sectional option-market pressure, and broad news climate. Discretionary variables describe call language, call structure, issuer-specific option state, issuer-specific news, and the event-day option reaction used only for next-day continuation tests. Table 3 lists the taxonomy. The separation is intentionally conservative: a feature is placed in the systematic block only when it is common across many issuers or aggregates the market; otherwise it is treated as discretionary.

Table 3. Feature taxonomy used before estimation.

Signal family	Variables	Block name	Economic intuition
Systematic	VIX_prev, vix_chg_5, vix_vol20	Market volatility state	Broad risk appetite before the call
Systematic	mkt_atm_iv_prev, mkt_atm_iv_chg_5, mkt_atm_iv_vol20	Market option-implied state	Common option repricing pressure
Systematic	mkt_iv_dispersion, mkt_contracts_log	Cross-sectional option activity	Breadth and intensity of common shocks
Systematic	mkt_news_count_7_log, mkt_news_sent_7, mkt_news_neg_7	Market news tone	News climate shared by many firms
Discretionary	txt_tone_balance, txt_positive_density, txt_negative_density, txt_uncertainty_density	Call language tone	Issuer-specific managerial language
Discretionary	txt_numeric_density, txt_forward_density, txt_qna_share, txt_qmark_log	Call structure and guidance	Information richness and analyst interaction
Discretionary	ATM_IV_prev, issuer_iv_chg_5, issuer_iv_chg_20, issuer_iv_rank60	Issuer option state	Firm-specific uncertainty before the event
Discretionary	issuer_put_call_oi_log, issuer_contracts_log, issuer_hv20,	Issuer option participation	Firm-specific option positioning and realized risk

	issuer_hv60		
Discretionary	issuer_news_count_7_log, issuer_news_sent_7, event_news_compound	Issuer news context	Firm-specific pre-call and event-day news flow
Discretionary reaction	event_abivchg, issuer_iv_chg_event, event_contracts_log	Event-day option reaction	Post-call interpretation available for continuation tests

Transcript features are constructed by deterministic text processing. The text is lowercased, tokenized, and scored with finance-oriented word lists for positive language, negative language, uncertainty, and forward-looking statements. Additional call-structure variables measure the logarithm of word count, the logarithm of line count, numeric density, question-mark intensity, Q&A share, and management-speaker share. These features are not meant to replace a large language model or a transformer embedding. Their purpose is to provide transparent, auditable proxies for the discretionary content that a future richer language model could expand.

Market-state variables are built from the options-IV panel and the news panel. The systematic option block includes lagged VIX, the five-day VIX change, 20-day VIX volatility, prior market average ATM IV, five-day market ATM-IV change, 20-day market ATM-IV volatility, market IV dispersion, and the logarithm of aggregate option contracts. The systematic news block includes seven-day market news count, average compound sentiment, and negative-sentiment share. Issuer-specific option variables include prior ATM IV, five-day and twenty-day issuer IV changes, the issuer's 60-day IV rank, put-call open-interest ratios, option contracts, and historical-volatility fields. Issuer-specific news variables measure seven-day news intensity and sentiment as well as event-day news tone.

Missing continuous features are imputed inside each modeling pipeline with the training-fold median. Logistic regression applies standardization after imputation and uses balanced class weights. Random forests and gradient boosting are estimated on the same feature blocks without standardization. All evaluation is chronological: event dates are sorted, expanding training windows are used, and each test fold occurs strictly after its training fold. The reported classification statistics are out-of-sample averages pooled across chronological folds. Accuracy, F1 score, AUC, Brier score, and the Matthews correlation coefficient are reported, with AUC emphasized because the task is primarily a probability-ranking problem.

Two additional diagnostics complement classification. First, the best next-day model is converted into an economic probability sort by ranking out-of-sample predictions into quintiles and reporting realized next-day abnormal IV changes. Second, robust OLS regressions summarize which variables retain directional explanatory power after the main rolling tests. The regressions are not the primary estimator; they are included to make the decomposition easier to interpret.

Results and Discussion

The matched sample is broader than the earlier small large-cap event panel. Table 4 lists the ten most frequently matched tickers. No single issuer dominates the final sample. The leading names contribute five or six matched calls each, and the final 563 events are spread across 254 tickers. This broader cross-section reduces the risk that the results reflect a small set of familiar technology or financial firms.

Table 4. Ticker coverage among the most frequently matched firms.

Ticker	Matched events	First event	Last event	Mean event abnormal IV change	Mean next-day abnormal IV change	Mean prior ATM IV
COO	6	2021-03-05	2022-05-04	-0.701	0.422	29.675
RMD	6	2021-04-28	2022-07-13	-0.013	0.364	31.200
IEX	5	2021-01-04	2022-01-04	-0.525	0.249	25.270
DXC	5	2021-09-30	2022-10-20	-0.547	-0.191	46.696
ITW	5	2021-02-08	2022-01-11	0.146	-0.431	22.882
APH	5	2021-01-26	2022-01-19	-0.014	0.154	25.092
LW	5	2021-07-19	2022-08-29	-0.306	-0.672	32.590
TDG	5	2021-04-30	2022-05-02	0.289	0.155	32.446
WEC	5	2021-02-11	2022-01-20	0.572	-0.916	20.416

UNM	5	2021-03-08	2022-01-04	1.564	-0.705	29.950
-----	---	------------	------------	-------	--------	--------

Table 5 reports descriptive statistics. Event-day abnormal IV changes have a mean close to zero and a standard deviation of 1.526 IV points. Next-day abnormal IV changes have a mean of 0.052 and a standard deviation of 1.362. The median prior ATM IV is 27.260, and the median prior VIX is 18.640. Transcript tone balance is positive on average, while uncertainty and forward-looking densities vary meaningfully across calls. This variation is important because the paper's central hypothesis requires issuer-level text heterogeneity rather than only market-wide volatility movement.

Table 5. Descriptive statistics for the matched earnings-call and option-IV sample.

Variable	Count	Mean	Std. dev.	Min	Median	Max
event_abivchg	563	-0.001	1.526	-22.504	0.005	12.608
next_abivchg	563	0.052	1.362	-7.323	0.030	13.917
two_day_abivchg	563	0.050	1.466	-8.747	0.052	12.937
ATM_IV_prev	563	28.530	6.805	14.990	27.260	57.600
issuer_iv_chg_20	563	-0.678	3.282	-28.800	-0.500	16.900
issuer_iv_rank60	563	0.419	0.287	0.000	0.401	1.000
txt_tone_balance	563	0.636	0.187	-0.009	0.667	1.000
txt_numeric_density	563	0.044	0.016	0.014	0.041	0.099
txt_uncertainty_density	563	0.011	0.005	0.000	0.010	0.028
txt_forward_density	563	0.014	0.006	0.002	0.013	0.036
VIX_prev	563	19.436	3.673	15.010	18.640	37.210
mkt_atm_iv_prev	563	40.956	2.588	38.027	40.362	54.626
mkt_iv_dispersion	563	22.268	1.192	19.757	21.818	27.369
issuer_news_sent_7	552	0.134	0.180	-0.421	0.095	0.831

Figure 2 plots annual event counts. Because the options-IV panel begins in late 2019 and the transcript files are concentrated in fiscal years 2020-2022, the usable event window runs from late 2020 through early 2023. The concentration is a feature of the release overlap rather than a sampling choice, and the chronological evaluation preserves the time ordering of those events.

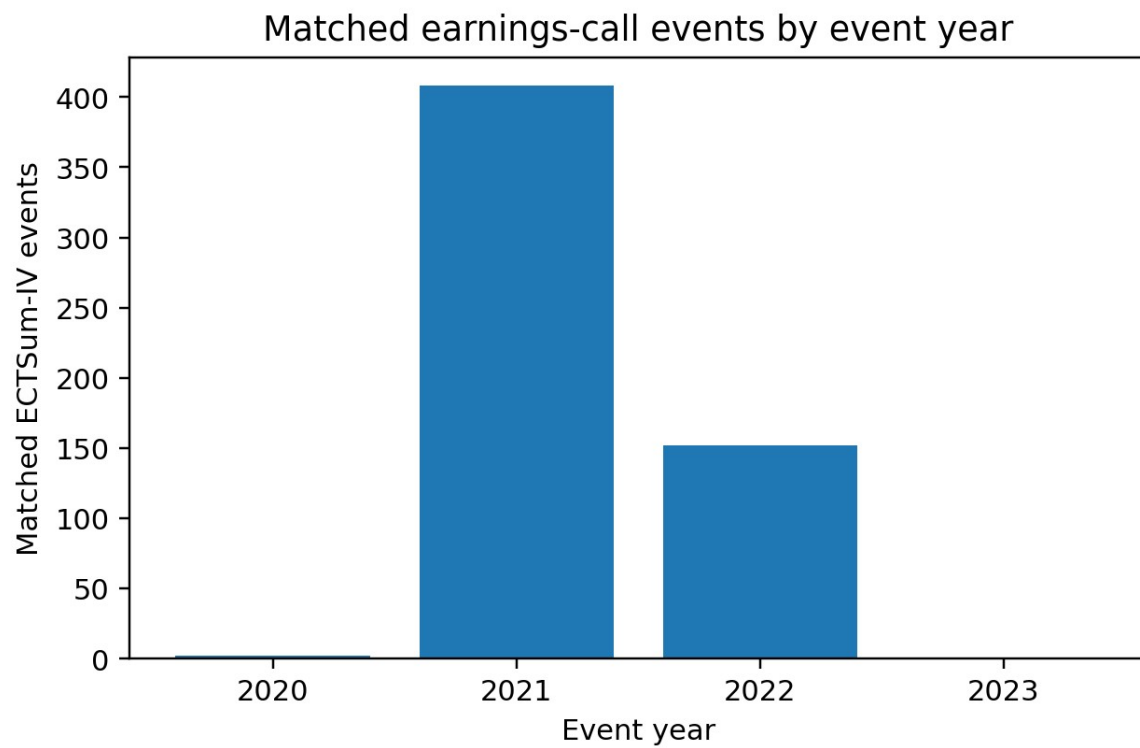


Figure 2. Annual event counts for raw transcripts and matched option-IV events.

The first descriptive relationship is between call tone and immediate option repricing. Figure 3 plots event-day abnormal IV changes against tone balance. The fitted slope is shallow, which is expected because option repricing is noisy and reflects many forms of uncertainty. Still, the relationship is not purely mechanical: text features carry some information, but the strongest signal emerges when call language is combined with issuer option state and news context.

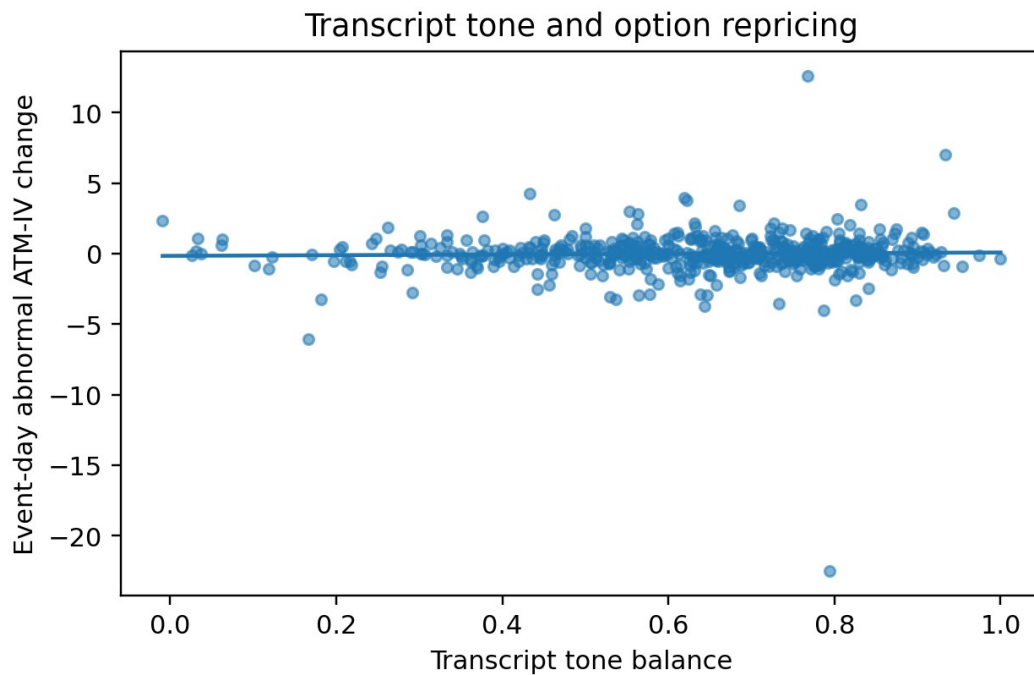


Figure 3. Event-day abnormal IV change and call-tone balance.

Figure 4 presents the same idea in quintile form. The tone-sort pattern is not monotone in every adjacent bin, but the upper part of the tone distribution shows more favorable abnormal-IV behavior than the bottom part. The figure supports using tone as one element of a discretionary block rather than as a stand-alone trading rule.

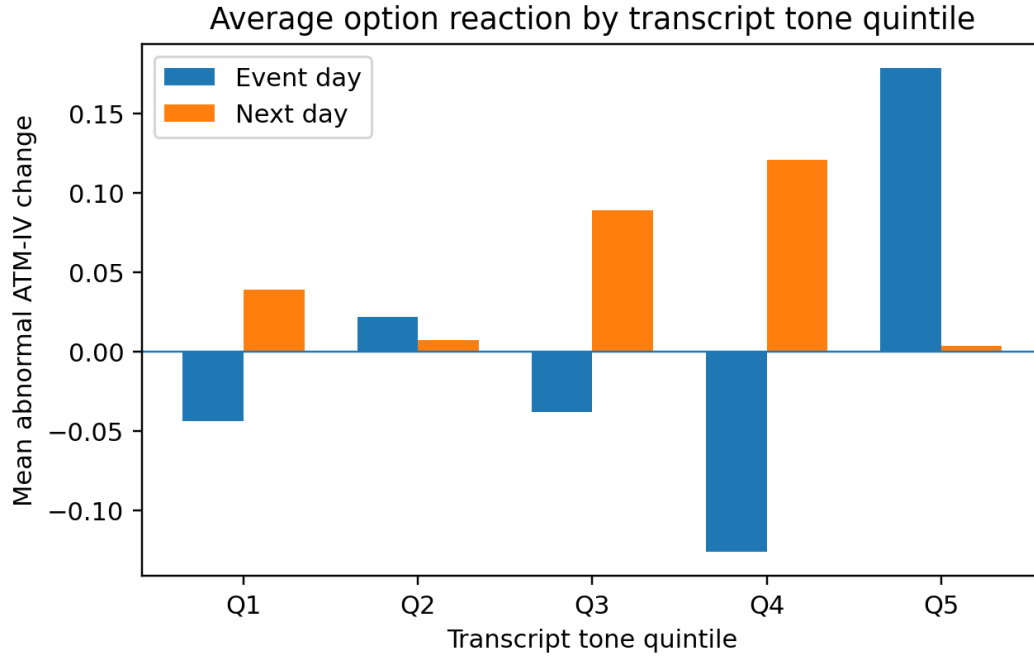


Figure 4. Mean abnormal IV changes across call-tone quintiles.

The event-day classification results in Table 6 are consistent with the paper's decomposition. Systematic variables alone are close to chance. The systematic logistic benchmark has an AUC of 0.499, and the systematic tree models are weaker. Transcript-only variables are better than systematic variables in the random-forest specification, but the gain remains limited. The strongest discretionary-pre models use both transcript features and issuer-specific option/news state, reaching AUC values around 0.557-0.565. The best event-day specification is the combined-pre logistic model, with accuracy of 0.567, F1 of 0.611, AUC of 0.579, and MCC of 0.135. The improvement is moderate, but it is aligned with the economic story: issuer-specific information creates the ranking, while market state adds context.

Table 6. Event-day abnormal-IV direction classification results.

Feature set	Model	N test	Accuracy	F1	AUC	Brier	MCC
Combined-pre	Logit	492	0.567	0.611	0.579	0.297	0.135
Discretionary-pre	Logit	492	0.559	0.565	0.565	0.278	0.118
Discretionary-pre	RF	492	0.557	0.602	0.560	0.249	0.114
Discretionary-pre	GB	492	0.575	0.609	0.557	0.253	0.150
Transcript-only	RF	492	0.510	0.507	0.528	0.252	0.021
Combined-pre	GB	492	0.530	0.563	0.520	0.262	0.060
Transcript-only	Logit	492	0.514	0.530	0.517	0.267	0.028
Systematic	Logit	492	0.504	0.523	0.499	0.289	0.007
Combined-pre	RF	492	0.494	0.503	0.498	0.260	-0.012
Transcript-only	GB	492	0.486	0.483	0.495	0.261	-0.028
Systematic	RF	492	0.463	0.450	0.462	0.273	-0.073
Systematic	GB	492	0.457	0.383	0.434	0.283	-0.085

Figure 5 displays representative event-day ROC curves. The combined-pre model sits above the systematic benchmark through much of the ranking range, while the discretionary-pre block remains competitive. The figure makes clear that the event-day gain is not driven by a single threshold choice; it appears in the probability ranking that AUC summarizes.

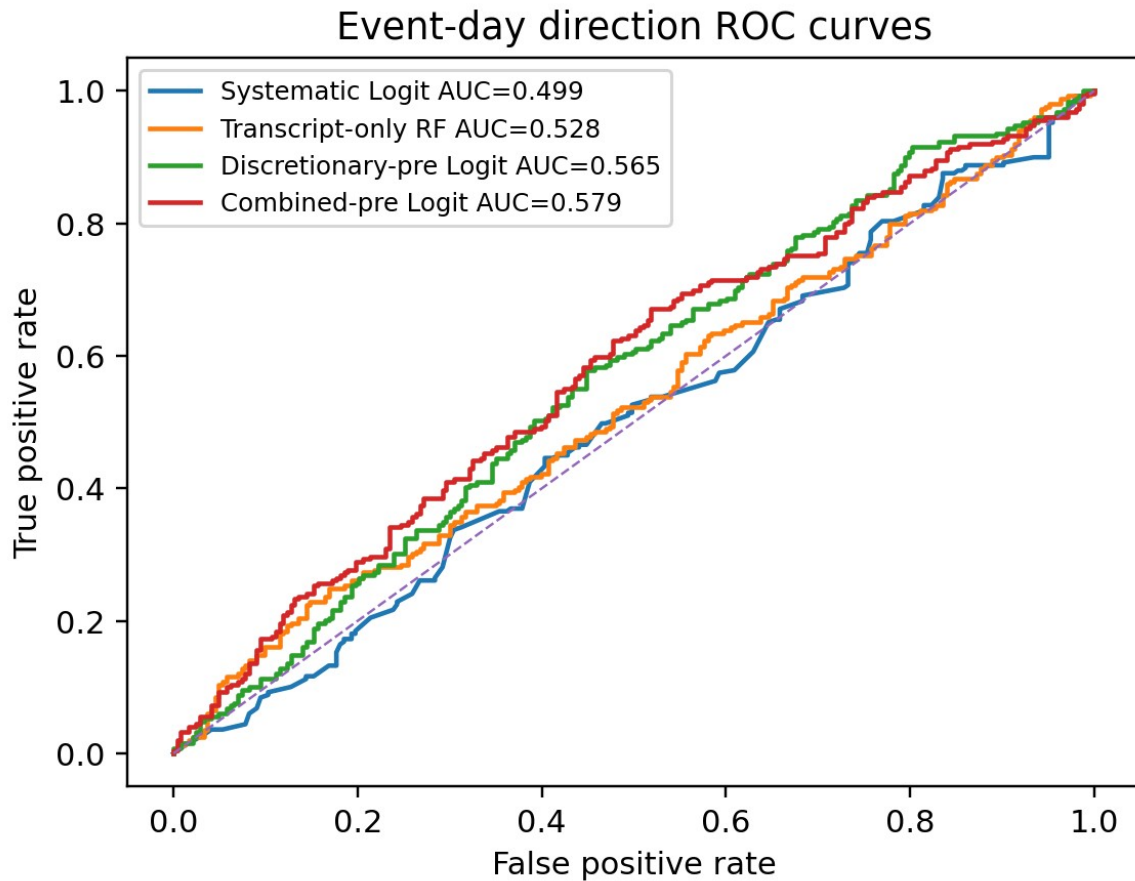


Figure 5. Event-day out-of-sample ROC curves for representative models.

The next-day continuation task is more difficult. Table 7 shows that the best model is the discretionary-plus-reaction logistic specification, with an AUC of 0.556 and accuracy of 0.535. Combined-all tree models are close, but adding systematic variables does not dominate the discretionary-plus-reaction block. This result is economically plausible. Once the call has occurred, event-day option reaction and issuer-specific state contain the most direct information about whether uncertainty repricing continues or reverses. Broad market variables help in some folds but do not replace the issuer-level signal.

Table 7. Next-day abnormal-IV direction classification results.

Feature set	Model	N test	Accuracy	F1	AUC	Brier	MCC
Discretionary+reaction	Logit	492	0.535	0.505	0.556	0.287	0.073
Combined-all	GB	492	0.533	0.527	0.549	0.254	0.066
Discretionary+reaction	GB	492	0.547	0.580	0.548	0.252	0.091
Discretionary+reaction	RF	492	0.522	0.503	0.544	0.252	0.047
Combined-all	RF	492	0.530	0.507	0.539	0.253	0.064
Discretionary-pre	RF	492	0.518	0.499	0.538	0.252	0.039
Transcript-only	GB	492	0.524	0.587	0.536	0.255	0.044
Transcript-only	RF	492	0.524	0.515	0.532	0.253	0.050
Combined-all	Logit	492	0.522	0.533	0.531	0.326	0.044
Transcript-only	Logit	492	0.498	0.489	0.521	0.259	-0.003
Discretionary-pre	Logit	492	0.508	0.481	0.512	0.292	0.020

Discretionary-pre	GB	492	0.504	0.536	0.495	0.259	0.005
Systematic	RF	492	0.488	0.417	0.488	0.258	-0.019
Systematic	GB	492	0.502	0.501	0.483	0.270	0.005
Systematic	Logit	492	0.459	0.434	0.477	0.291	-0.079

Figure 6 shows the corresponding next-day ROC curves. The curves are closer together than in the event-day task, which is consistent with the lower AUCs. Even so, the best discretionary-plus-reaction specification remains above the systematic-only benchmark, reinforcing the interpretation that continuation is mainly an issuer-specific phenomenon.

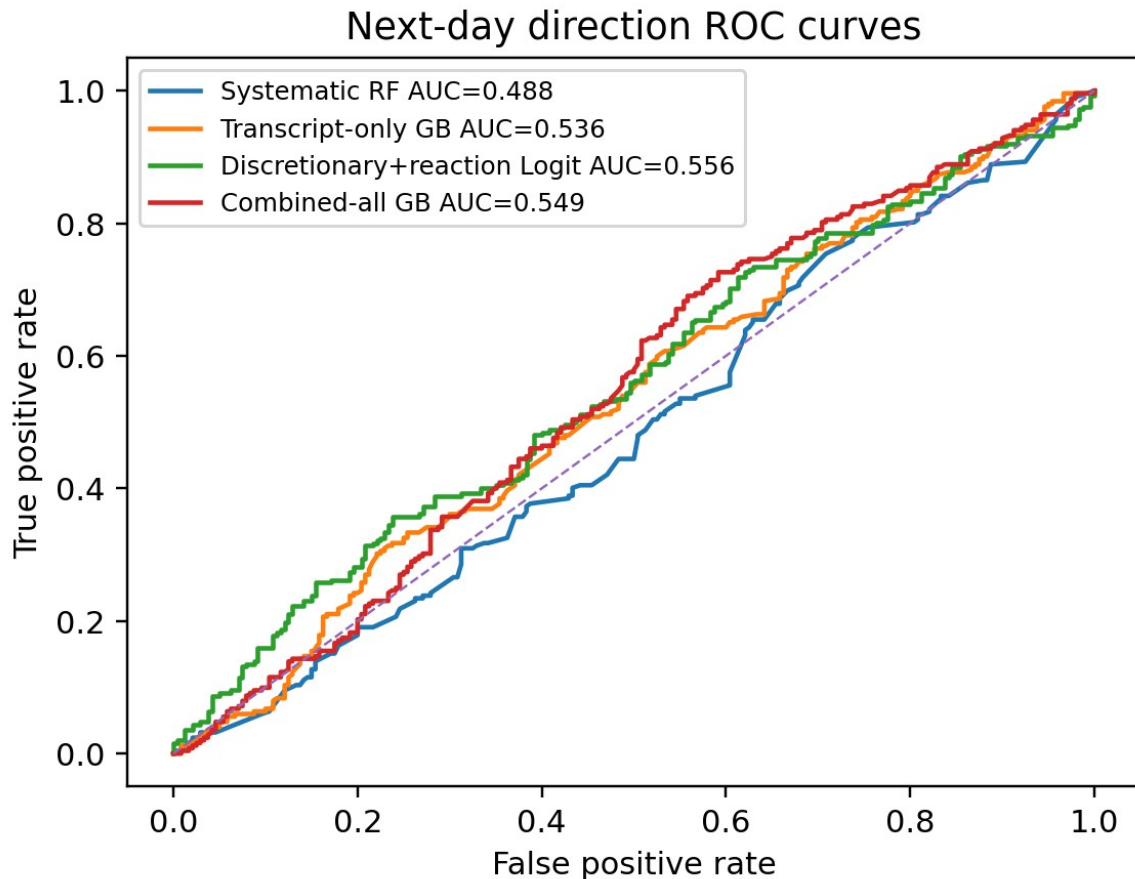


Figure 6. Next-day out-of-sample ROC curves for representative models.

The economic ranking diagnostic in Table 8 provides a clearer view of the next-day signal. The bottom probability quintile has a mean next-day abnormal IV change of -0.168, while the top probability quintile has a mean of 0.385. The top-minus-bottom spread is 0.552 IV points with a Welch t-statistic of 2.47 and a p-value of 0.0146. The hit rate rises from 45.45 percent in Q1 to 62.63 percent in Q5. This is the strongest evidence in the paper: even though classification AUC is modest, the probability ranking separates low- and high-continuation events in economically meaningful terms.

Table 8. Out-of-sample next-day probability quintile sorts.

Probability quintile	N	Mean next-day abnormal IV change	Median next-day abnormal IV change	Next-day IV-up rate	Mean predicted probability
Q1	99	-0.168	-0.107	45.45%	0.115
Q2	98	0.008	-0.023	47.96%	0.320
Q3	98	0.013	0.055	55.10%	0.459
Q4	98	-0.037	-0.105	44.90%	0.592
Q5	99	0.385	0.129	62.63%	0.799

Figure 7 reports permutation importance for the next-day combined random forest. The most important variables are event-day issuer IV change, event-day abnormal IV change, event-day option contracts, transcript length, and market ATM-IV volatility. This ranking is informative. It shows that the model does not rely only on broad market volatility; it also uses the issuer's own event-day option repricing and the information scale of the call.

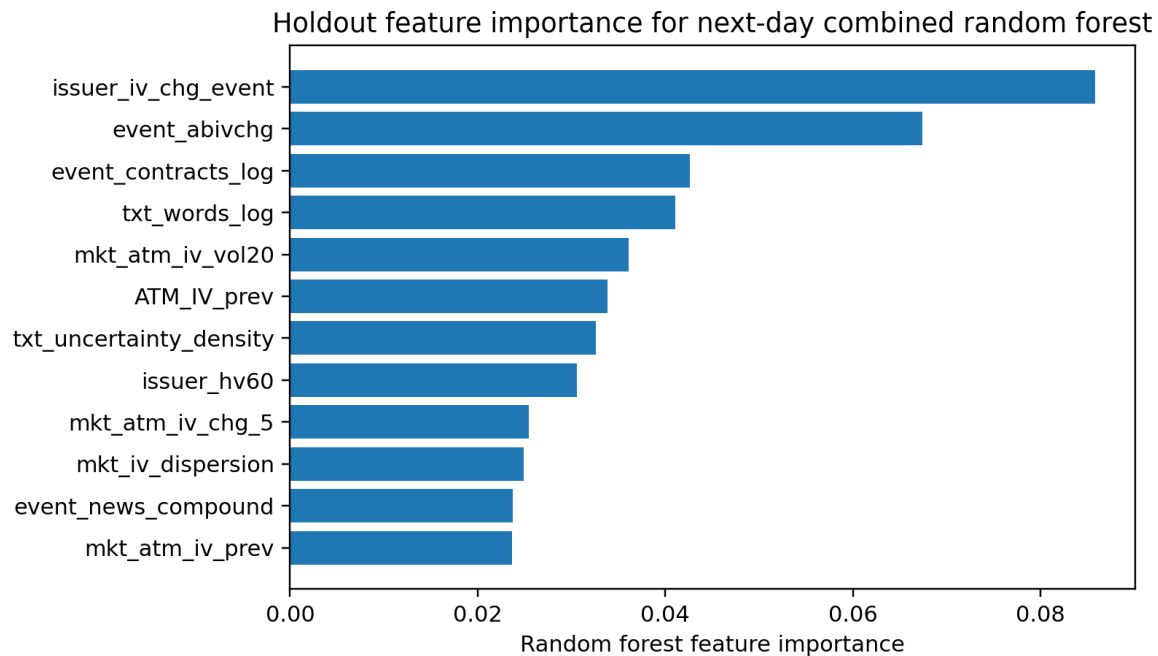


Figure 7. Permutation importances for the next-day combined random forest.

Figure 8 visualizes the quintile sort from Table 8. The most important comparison is the separation between Q1 and Q5 rather than perfect monotonicity across all adjacent quintiles. In a 492-observation out-of-sample panel, local non-monotonicity is expected. The upper tail still identifies calls after which implied-volatility adjustment is materially more likely to continue.

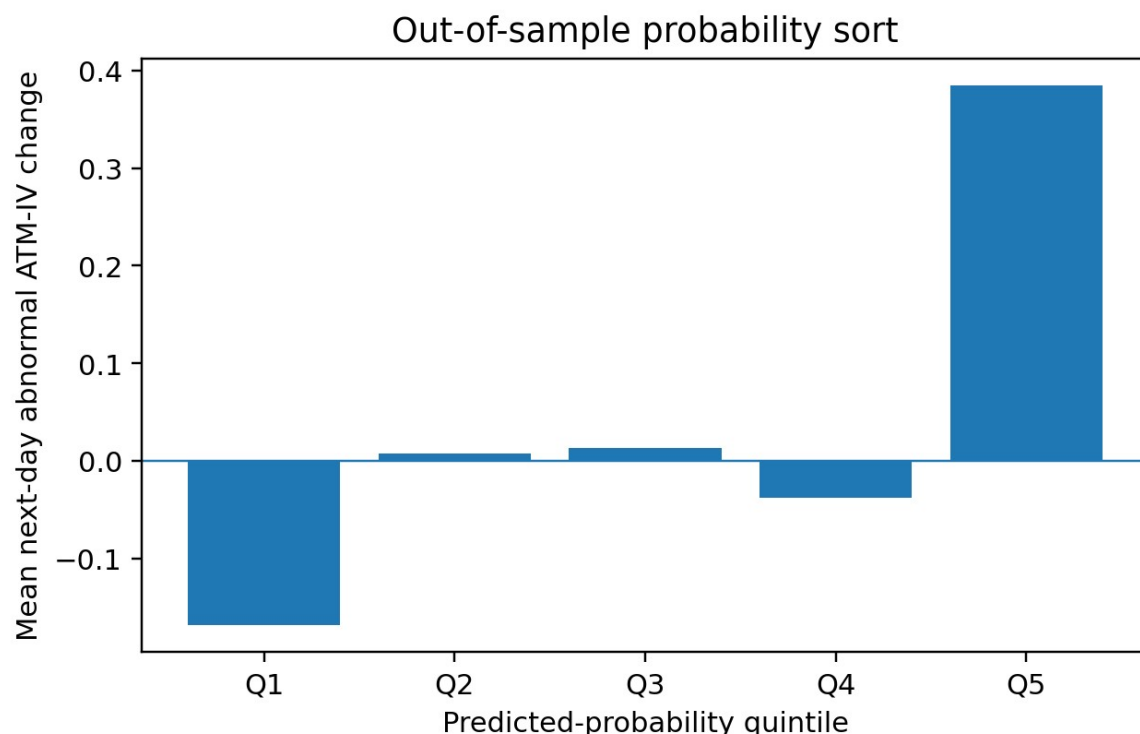


Figure 8. Mean next-day abnormal IV change across out-of-sample probability quintiles.

The OLS decomposition in Table 9 supports a cautious version of the same story. For event-day abnormal IV change, issuer 60-day IV rank is negative and significant at the 1 percent level, while numeric density is negative at the 5 percent level. Prior ATM IV, issuer 20-day IV change, and prior VIX are marginally significant. For next-day abnormal IV change, prior VIX is positive and significant at the 5 percent level, while event-day abnormal IV change is negative and significant at the 5 percent level. The signs indicate partial reversal after large event-day IV moves and stronger continuation pressure when broad market volatility is already elevated.

Table 9. OLS coefficient summary with robust standard errors.

Variable	Event-day coef.	Event-day p	Next-day coef.	Next-day p
Prior ATM IV	-0.094	<0.10	-0.131	<0.10
Prior VIX	0.121	<0.10	0.138	<0.05
Event-day abnormal IV change			-0.289	<0.05
Event-day contracts log			-0.015	n.s.
Issuer 20-day IV change	-0.103	<0.10	-0.062	n.s.
Issuer event-day IV change			-0.067	n.s.
Issuer 60-day IV rank	-0.228	<0.01	-0.079	n.s.
Issuer 7-day news sentiment	-0.017	n.s.	-0.003	n.s.
Put-call open-interest log ratio	-0.005	n.s.	-0.040	n.s.
Market 5-day ATM-IV change	0.031	n.s.	0.027	n.s.
Market IV dispersion	-0.090	n.s.	-0.078	n.s.
Forward-looking density	-0.025	n.s.	0.034	n.s.
Numeric density	-0.080	<0.05	-0.006	n.s.
Call tone balance	0.017	n.s.	0.018	n.s.
Uncertainty density	-0.025	n.s.	0.041	n.s.

Table 10 examines state dependence by splitting the sample at the median of prior VIX and transcript text intensity. The regime table is less sharp than the probability sort, but it still helps interpret the model results. High-VIX events do not automatically produce stronger continuation; rather, continuation depends on how market state interacts with issuer-specific call information and event-day option activity. This is why systematic variables are better understood as conditioning variables than as stand-alone signals.

Table 10. State dependence by VIX and transcript-intensity regimes.

VIX regime	Text-intensity regime	N	Mean event abnormal IV change	Mean next-day abnormal IV change	Event IV-up rate	Next-day IV-up rate
High	High	147	0.039	0.003	47.62%	48.30%
High	Low	142	0.074	0.131	54.93%	54.93%
Low	High	135	-0.021	0.040	45.93%	48.89%
Low	Low	139	-0.102	0.034	51.80%	54.68%

A final chronological holdout is reported in Table 11. Training on the earlier portion of the matched panel and testing on the final block gives weaker performance for the combined random forest than the rolling-fold average. This result narrows the interpretation of the paper. The evidence is not a claim that every estimator remains stable in every final-period split. The stronger and more robust result is the rolling out-of-sample ranking and the top-versus-bottom probability spread. That is the appropriate level of inference for a 563-event transcript-option panel.

Table 11. Final chronological holdout for the next-day combined random forest.

Feature set	Model	N test	Accuracy	F1	AUC	Brier	MCC
Combined-all	RF	51	0.412	0.167	0.432	0.274	-0.065

Taken together, the evidence supports a layered interpretation. Discretionary variables create most of the predictive structure around earnings calls. Systematic variables matter, but mainly as context. They help describe whether the broader volatility environment is receptive to continued uncertainty repricing, but they do not by themselves explain event-day or next-day option-implied adjustment. The data therefore align the empirical design with the paper's original motivation: a transcript-based earnings-call system should enrich the discretionary block with language and interaction features while keeping market-regime variables separate.

Limitations

The paper has four practical limitations. First, the sample is constrained by the overlap between ECTSum, the S&P 500 news panel, and the options-IV panel. The final matched sample is broad in issuer count but concentrated in the 2020-2023 event window. Second, the outcome is option-implied volatility rather than stock-return abnormal return. This is a deliberate choice because the options panel provides the cleanest overlap with the transcript years, but it means the results speak to uncertainty repricing rather than to equity return drift. Third, the text variables are transparent lexicon and structure features rather than large transformer embeddings. That choice supports auditability, but richer embeddings may capture semantic novelty and question-answer asymmetry more completely. Fourth, the holdout result is weaker than the rolling probability sort, so the empirical claim should remain modest: the data show an economically ordered transcript-option signal, not a uniformly large or frictionless trading strategy.

These limits define the scope of the findings. Within the data, the paper directly observes earnings-call text, issuer-specific option state, event-day option reaction, market volatility, and news climate. The decomposition therefore answers the intended discretionary-versus-systematic question more directly than a date-only earnings-event panel. A natural extension would add intraday call timestamps, more granular option maturities, and transformer-based transcript representations while preserving the same pre-estimation feature taxonomy.

Conclusion

This paper provides a transcript-based empirical answer to the central question: earnings-call trading signals separate into discretionary and systematic components, and the discretionary component carries most of the observed predictive structure. On 563 matched earnings-call events from 254 firms, systematic variables alone are close to chance for event-day abnormal-IV direction, while discretionary variables improve the ranking and combined models perform best on the event day. For next-day continuation, the strongest model is a discretionary-plus-reaction specification. The top probability quintile earns a 0.552-point higher next-day abnormal IV change than the bottom quintile. The practical implication is a two-layer design. The first layer should identify the firm-specific call signal and issuer option state. The second layer should scale or filter that signal by market-wide volatility and option-market conditions. This decomposition provides a disciplined foundation for future LLM-based earnings-call systems because it prevents language features from being confused with common market state.

Data Availability

The data releases used for the study are publicly accessible from the following sources:

ECTSum repository ZIP: <https://github.com/rajdeep345/ECTSum/archive/refs/heads/main.zip>

ECTSum data directory: <https://github.com/rajdeep345/ECTSum/tree/main/data/final>

S&P 500 news, ratios, and price dataset page: https://huggingface.co/datasets/pmoe7/SP_500_Stocks_Data-ratios_news_price_10_yrs

S&P 500 daily ratios ZIP: https://huggingface.co/datasets/pmoe7/SP_500_Stocks_Data-ratios_news_price_10_yrs/resolve/main/sp500_daily_ratios_20yrs.zip

S&P 500 news CSV: https://huggingface.co/datasets/pmoe7/SP_500_Stocks_Data-ratios_news_price_10_yrs/resolve/main/sp500_news_290k_articles.csv

S&P 500 options-IV dataset page: <https://huggingface.co/datasets/gauss314/options-IV-SP500>

S&P 500 options-IV CSV: https://huggingface.co/datasets/gauss314/options-IV-SP500/resolve/main/data_IV_USA.csv

References

- [1] R. Ball and P. Brown, "An empirical evaluation of accounting income numbers," *Journal of Accounting Research*, vol. 6, no. 2, pp. 159-178, 1968.
- [2] W. H. Beaver, "The information content of annual earnings announcements," *Journal of Accounting Research*, vol. 6, pp. 67-92, 1968.
- [3] Yuanzheng Chen, Yitian Zhang, and Matt Sherman, "Going Concern and Bankruptcy Prediction under Extreme Class Imbalance: Cost-Sensitive Learning, Resampling, and Focal Loss with Explainable Financial-Ratio Portraits", *JACS*, vol. 4, no. 4, pp. 80–96, Apr. 2024, doi: 10.69987/JACS.2024.40407.
- [4] V. L. Bernard and J. K. Thomas, "Evidence that stock prices do not fully reflect the implications of current earnings for future earnings," *Journal of Accounting and Economics*, vol. 13, no. 4, pp. 305-340, 1990.
- [5] L. K. C. Chan, N. Jegadeesh, and J. Lakonishok, "Momentum strategies," *Journal of Finance*, vol. 51, no. 5, pp. 1681-1713, 1996.
- [6] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *Journal of Finance*, vol. 48, no. 1, pp. 65-91, 1993.
- [7] S. P. Kothari, "Capital markets research in accounting," *Journal of Accounting and Economics*, vol. 31, no. 1-3, pp. 105-231, 2001.
- [8] Jiaying Jin, Tina Huang, and Sam Lu, "A Model-Risk-Friendly Probability of Default Workflow: Calibration, Distribution-Free Uncertainty Quantification, and SHAP Explanations on the UCI Credit Card Default Dataset", *JACS*, vol. 4, no. 6, pp. 74–85, Jun. 2024, doi: 10.69987/JACS.2024.40606.
- [9] F. Li, "The information content of forward-looking statements in corporate filings-A naive Bayesian machine learning approach," *Journal of Accounting Research*, vol. 48, no. 5, pp. 1049-1102, 2010.
- [10] T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *Journal of Finance*, vol. 66, no. 1, pp. 35-65, 2011.
- [11] D. A. Matsumoto, M. Pronk, and E. Roelofsen, "What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions," *The Accounting Review*, vol. 86, no. 4, pp. 1383-1414, 2011.
- [12] A. K. Davis, J. A. Piger, and L. M. Sedor, "Beyond the numbers: Measuring the information content of earnings press release language," *Contemporary Accounting Research*, vol. 29, no. 3, pp. 845-868, 2012.
- [13] D. F. Larcker and A. A. Zakolyukina, "Detecting deceptive discussions in conference calls," *Journal of Accounting Research*, vol. 50, no. 2, pp. 495-540, 2012.
- [14] W. J. Mayew and M. Venkatachalam, "The power of voice: Managerial affective states and future firm performance," *Journal of Finance*, vol. 67, no. 1, pp. 1-43, 2012.
- [15] B. J. Bushee, D. A. Matsumoto, and G. S. Miller, "Managerial and investor responses to disclosure regulation: The case of Reg FD and conference calls," *The Accounting Review*, vol. 79, no. 3, pp. 617-643, 2004.
- [16] X. Huang, S. H. Teoh, and Y. Zhang, "Tone management," *The Accounting Review*, vol. 89, no. 3, pp. 1083-1113, 2014.
- [17] S. J. Brown and J. B. Warner, "Using daily stock returns: The case of event studies," *Journal of Financial Economics*, vol. 14, no. 1, pp. 3-31, 1985.
- [18] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654-669, 2018.
- [19] B. T. Kelly, S. Pruitt, and Y. Su, "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics*, vol. 134, no. 3, pp. 501-524, 2019.
- [20] Jiaying Jin, Tina Huang, and Sam Lu, "Cost-Sensitive Learning, Simulated PU Learning, and One-Class Autoencoding for Extreme-Imbalance Credit Card Fraud Detection", *JACS*, vol. 4, no. 6, pp. 64–73, Jun. 2024, doi: 10.69987/JACS.2024.40605.
- [21] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems* 30, 2017, pp. 5998-6008.
- [22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171-4186.
- [23] D. Araci, "FinBERT: Financial sentiment analysis with pre-trained language models," *arXiv:1908.10063*, 2019.
- [24] T. Brown et al., "Language models are few-shot learners," in *Advances in Neural Information Processing Systems* 33, 2020, pp. 1877-1901.
- [25] Yuanzheng Chen, Yitian Zhang, David Chau, and Matt Sherman, "Credit Card Default Risk Tiering with Probability Calibration and Uncertainty-Driven Rejection: A Reproducible Study on the UCI Credit Card Clients Dataset", *JACS*, vol. 3, no. 4, pp. 31–47, Apr. 2023, doi: 10.69987/JACS.2023.30403.
- [26] pmoe7, "SP_500_Stocks_Data-ratios_news_price_10_yrs," Hugging Face dataset, 2023.
- [27] gauss314, "options-IV-SP500," Hugging Face dataset, 2023.