AP 2026-002

Do Podcasts Move the Stock Market ?

Marten Laudi, Janik Ole Wecks

Academic paper

Netspar

Do Podcasts Move the Stock Market?

Marten Laudi* Janik Ole Wecks†

This version: April 2026

Abstract

We construct a large dataset of transcribed investing podcast episodes and identify over 100,000 time-stamped discussions on individual stocks to study how this coverage relates to trading activity and investor outcomes. We find that podcast discussions are associated with significant increases in stock-level turnover, retail trading among Robinhood users, and absolute returns with effects that are largest on the day of release and persist on subsequent days. Responses are larger for podcasts with greater audience reach and for discussions that are forward-looking or contain explicit buy or sell recommendations. Moreover, investor reactions to earnings announcements are substantially amplified when these events are accompanied by podcast coverage, consistent with podcasts facilitating the diffusion and interpretation of publicly available information. In line with this mechanism, trading responses are most pronounced among retail investors, who face higher information frictions. We then assess whether podcast-driven trading is beneficial to investors and find no evidence that trading strategies based on podcast content consistently outperform passive benchmarks. Instead, such strategies often underperform at realistic implementation horizons. Taken together, our results suggest that investing podcasts accelerate the incorporation of publicly available information into prices rather than providing an exploitable information advantage.

JEL Classification: G14, G40

Keywords: Media and Financial Markets; Information Diffusion; Behavioral Finance; Retail Investors; Podcasts

*KLU Hamburg. marten.laudi@klu.org

†Corresponding author. University of Bremen. wecks@uni-bremen.de

We gratefully acknowledge financial support from Behavioral Finance e.V. We also thank the Data Science Center of the University of Bremen for providing high-performance computing resources and Listen Notes for generously supplying data used in this research. Part of this research was conducted while Janik Ole Wecks was visiting researcher at Concordia University Montreal.

1 Introduction

Investing podcasts, in which hosts discuss individual stocks, interpret news, and offer investment advice, have become a widely used source of financial information. Episodes are released frequently and reach large audiences, with more than half of the U.S. population listening to podcasts (Marshall et al., 2025). Yet it remains unclear whether podcast discussions about individual stocks influence trading behavior or affect investor outcomes.

In this paper, we construct a novel dataset of more than 25,000 transcribed podcast episodes that captures time-stamped discussions of individual stocks, which we classify based on their sentiment, wording, expressions, and the presence of explicit investment recommendations. We exploit the timing and content of these discussions to link podcast coverage to stock-level trading activity, investor holdings, and the portfolio performance of investors trading on podcast content.

Podcast discussions have the potential to influence investor behavior by expanding the reach and reducing processing costs of public information. Podcasts provide repeated and easily accessible discussions of individual firms, often accompanied by explicit interpretations and investment recommendations, which can make existing information more salient and easier to act upon. By lowering the cost of accessing and acting on public information, such discussions can affect how and when investors trade.

Our results show that podcast coverage is followed by significant increases in stock-level turnover, absolute returns, and changes in retail investor stock holdings, with effects that are strongest on the day of release and remain positive in the following days. The sharp timing of the response, the absence of pre-trends in turnover and absolute return, and null results in placebo tests make it unlikely that these patterns reflect spurious correlations. The findings are robust across alternative exposure measures, specifications, and sample restrictions.

A range of patterns suggests that these effects reflect the influence of podcast coverage rather than common-news shocks. First, we exploit variation in podcast release schedules. Many shows follow a fixed weekly cadence, so episode timing is determined by production schedules rather than by contemporaneous news. We find that trading responses remain economically and statistically significant for these regularly scheduled episodes. Second, we consider heterogeneity in trading responses and find stronger effects for podcasts with greater audience reach. Responses are also more pronounced when discussions are forward-looking and include explicit

investment recommendations. In addition, negative sentiment of discussions is associated with stronger trading responses than positive sentiment, in line with evidence of negativity bias in financial media and market outcomes (Tetlock, 2007). Taken together, these patterns indicate that trading responses are driven by podcast coverage itself rather than by coincident news.

Consistent with podcasts acting as information intermediaries, investor responses to earnings announcements and M&A announcements are substantially stronger when these events are accompanied by podcast coverage, suggesting that podcasts amplify the diffusion of publicly available information.

We next examine the implications of these trading responses for investor portfolios. Prior research shows that media-driven retail trading can be harmful to portfolio outcomes (Barber & Odean, 2000, 2008). If podcast discussions lower the cost of accessing and acting on public information, the induced trading may not improve investment outcomes. By the time investors trade, the information may already be reflected in prices. To assess this, we construct trading strategies that reflect realistic ways in which investors might trade on podcast content.

We find that none of the podcast-based strategies outperform passive benchmarks. In particular, strategies that build positions based on repeated exposure or take contrarian positions yield negative returns, and a long–short strategy based on podcast content does not generate positive returns at realistic implementation horizons. These findings indicate that podcast-driven trading does not provide an exploitable advantage. Instead, podcasts appear to increase trading activity without improving outcomes.

Taken together, our results suggest that investing podcasts influence trading behavior but do not improve investor outcomes. Although podcast discussions generate immediate trading responses, this additional activity does not translate into outperformance of passive benchmarks. This pattern suggests that podcasts accelerate the diffusion of publicly available information and its incorporation into prices, so that it is already reflected in prices by the time most listeners trade.

These findings contribute to the literature on media and financial markets by identifying podcasts as an economically relevant channel through which public information is disseminated and incorporated into trading activity and prices. Prior research shows that traditional media and online platforms can influence trading by transmitting and aggregating information (Antweiler & Frank, 2004; Chen et al., 2014; Tetlock, 2007). Engelberg and Parsons (2011) show that differences in local media coverage of the same earnings announcement lead to differences in

local trading activity. Related work shows that making information more accessible increases investor attention and trading activity, particularly among individual investors (Blankespoor et al., 2019). Our results highlight podcasts as a distinct and increasingly important medium through which public information reaches investors. We show that podcast coverage amplifies investor responses to news and primarily affects retail investors, consistent with podcasts accelerating the diffusion of publicly available information. This interpretation is in line with recent evidence that the broader diffusion of public information across investors strengthens market reactions. For example, Hirshleifer et al. (2025) show that earnings announcements originating in more socially connected regions generate stronger and faster price and trading responses. More generally, prior work shows that the spread of information across investors affects trading behavior, and that when attention is limited, the salience of information influences how it is incorporated into prices (Hirshleifer & Teoh, 2003; Hong et al., 2004, 2005). We add to this literature by identifying podcasts as a media channel that expands the reach and salience of public information, thereby facilitating its incorporation into trading activity and prices.

The paper also relates to literature on media-driven trading (Barber & Odean, 2008; Focke et al., 2020). Individual investors often trade to their own detriment when responding to salient but widely available information (Barber & Odean, 2000, 2013). Consistent with this, we find that trading strategies that approximate how investors might act on podcast content generate increased trading activity without improving portfolio performance.

Our results have direct implications for retail investors. As podcasts become an increasingly prominent source of financial information, they can stimulate trading activity without improving portfolio performance. Strategies that approximate how listeners might trade, such as gradually accumulating positions from repeated exposure or overweighting positively discussed stocks, do not outperform passive benchmarks. Taken together, these findings suggest that podcast-driven trading may increase activity and associated costs without corresponding gains, with potential consequences for investor welfare.

2 Data and Sample Selection

2.1 Podcast Data

We construct a dataset of stock-specific discussions from investing podcasts. We obtain podcast metadata from Listen Notes,[1] a large podcast search engine and data provider that aggregates information on podcast shows, including language, category tags, and popularity measures. Using this source, we identify all English-language podcasts tagged as "Investing". Because this category is broad and includes content on financial education, passive investing, real estate, and lifestyle, we manually screen the 1,000 most popular podcasts with the investing tag.[2] We apply two selection criteria. First, the podcast must focus on equity investing. Second, it must regularly discuss individual stocks rather than focusing on broad economic and market developments. This procedure yields a final sample of 93 podcast shows (Table A.1 lists all podcasts in the sample).[3]

For these podcasts, we collect episode audio files from their RSS feeds as indexed by Listen Notes. We transcribe all available episodes using OpenAI Whisper. The resulting corpus comprises 25,350 episodes from January 2017 through December 2021. We focus on this period to center the sample on the Robintrack window (May 2018–August 2020), which provides data on stock-level holdings of retail investors on Robinhood (see Section 2.3), while extending it by one year on each side to facilitate lead–lag tests and event-window analyses.

Episode transcripts cover a wide range of topics beyond individual stocks. To isolate stock-level content, we process all transcripts with Llama-3.1-8B-Instruct, which identifies passages in which hosts discuss a specific firm or stock. We extract each passage as a self-contained text snippet and link it to a ticker and a calendar date. This procedure transforms the episode-level corpus into a stock-level panel in which the unit of observation is a text snippet about a given stock on a given day.

The resulting panel contains 102,623 stock-specific snippets covering 4,570 distinct tickers across 1,650 calendar days. Snippets contain 147 words on average (median: 111), with a total text volume of about 15 million words. About 70% of episodes yield at least one stock-related snippet. The remaining 30% consist of special episodes, Q&A sessions, or broad discussions

[1] listennotes.com

[2] Screening is based on the podcast description and a review of both the most recent episode and a randomly selected earlier episode.

[3] One podcast (*Motley Fool Money*) appears under two separate RSS feeds, yielding 94 unique podcast IDs in the data.

without references to individual stocks. Coverage is concentrated in large firms, well-known to consumers. Tesla (5,778 snippets), Apple (3,692), and Amazon (3,267) are the most frequently discussed stocks, consistent with a retail-oriented podcast sample. The top 10 tickers account for 22.2% of all snippets, and the top 50 account for 45.2%.[4]

2.2 Podcast Content Measures

For each snippet, we measure several conceptually distinct dimensions of the discussion content, including its sentiment, the degree to which it is forward-looking, how closely it resembles an actionable investment recommendation, and the conviction with which the speaker expresses a view. To construct these measures, we combine three broad approaches that vary in complexity and transparency. First, we use large language models (LLMs), which we instruct via standardized prompts to read and score each snippet. Second, we use pretrained classifiers and text-similarity methods that compare the wording of a snippet to predefined reference texts. Third, we use simple keyword counts based on established financial dictionaries. Using multiple methods for each dimension ensures that our results are not driven by the idiosyncrasies of any single approach. Detailed variable definitions, prompt designs, and construction procedures for all measures are reported in Table OA.1 in the Online Appendix.

Our primary sentiment measure is a continuous LLM sentiment score on the interval $[-1, +1]$, where -1 denotes strongly bearish language and $+1$ denotes strongly bullish language. To obtain this score, we apply a standardized prompt to each individual snippet, instructing the model to assess the directional tone of the stock discussion and return a numerical rating (see Table OA.1 in the Online Appendix for the exact prompt design and variable definitions). We use a prompt-based approach because podcast transcripts are conversational and informal in register, which differs substantially from the written financial text (e.g., SEC filings or analyst reports) on which domain-specific models such as FinBERT are trained. The LLM approach allows us to tailor the classification task to our specific context of stock-level discussions in spoken financial media and to capture nuances in tone that pretrained classifiers may miss when applied outside their training domain (e.g., irony, jokes, sarcasm, or colloquial emphasis). To verify that our results do not depend on the choice of sentiment measure, we construct several alternatives. First, we use FinBERT, a finance-adapted language model developed for extracting informa-

[4] We address this concentration later by controlling for firm size and by conducting robustness tests that exclude the most heavily covered firms.

tion from financial text (Huang et al., 2023). Operationally, we transform the model output into a continuous sentiment score by taking the difference between the positive and negative class probabilities. Second, we use the Loughran–McDonald dictionary, which is the standard domain-specific lexicon for financial sentiment and is designed to avoid the misclassification problems of general-purpose sentiment dictionaries in financial text (Loughran & McDonald, 2011). Third, we construct a prototype-based sentiment score using term frequency–inverse document frequency (TF-IDF) weighting. TF-IDF converts each snippet into a numerical vector that reflects which words are used frequently in the snippet but rarely in the corpus as a whole, thereby emphasizing distinctive vocabulary. We then measure the similarity of each snippet's word vector to predefined positive and negative reference vectors (prototypes), following the general vector-space logic of term-weighting approaches in text analysis (Salton & Buckley, 1988). A snippet that uses language closer to the positive prototype receives a higher sentiment score, and vice versa (see Table OA.1 in the Online Appendix for the prototype construction). Correlations across methods are moderate (FinBERT–LLM: $\rho = 0.42$; FinBERT–LM Dict.: $\rho = 0.40$), which suggests that these measures capture related but not identical aspects of sentiment.[5]

We measure the extent to which podcast discussions are forward-looking, that is, whether hosts make statements about future expectations, forecasts, or anticipated developments rather than merely describing past events. Our primary measure is a continuous LLM forward-looking score on $[0, 1]$, obtained by applying a standardized prompt to each snippet, analogous to the sentiment scoring described above, that instructs the model to rate how much of the discussion is oriented toward the future (see Table OA.1 in the Online Appendix for details). To verify robustness, we employ two alternatives. First, we use FinBERT-FLS, a pretrained classifier designed to detect forward-looking statements (FLS) in corporate disclosures. The model categorizes text into three classes: *Specific FLS* (e.g., quantified earnings guidance), *Non-Specific FLS* (e.g., qualitative outlook statements), and *Not-FLS* (backward-looking or factual content). It is trained on 3,600 manually labeled sentences from the Management Discussion and Analysis (MD&A) sections of Russell 3000 annual reports (Bozanic et al., 2018; Huang, 2022; Muslu et al., 2015). We aggregate the probabilities of the two forward-looking classes into a single continuous score. Second, we count forward-looking keywords using word lists developed by Li

[5]Figures OA.1 and OA.2 in the Online Appendix plot the time-series behavior of all sentiment and content measures over the sample period.

(2010) for identifying forward-looking language in corporate filings.

We measure actionability, defined as the extent to which a snippet resembles an explicit investment recommendation rather than a purely descriptive discussion. Our primary actionability measure uses the TF-IDF similarity approach described above: we compare each snippet's word vector to four reference profiles that represent typical *buy*, *sell*, *hold*, and *none* (i.e., no recommendation) language. Actionability is then defined as the degree to which the snippet's wording resembles any of the three recommendation categories rather than the non-recommendation reference (see Table OA.1 in the Online Appendix for the prototype construction and scoring procedure). This makes actionability conceptually distinct from sentiment: a snippet can be positive in tone without telling the listener what to do, and a negative snippet can remain purely descriptive. Consistent with this distinction, actionability is largely orthogonal to sentiment in our data ($\rho < 0.12$). As a supplementary measure, we count explicit buy and sell keywords using dictionary-based word lists.

We measure explicit recommendation language separately. Using a dedicated LLM prompt, we instruct the model to determine whether a snippet contains a concrete investment recommendation and, if so, to classify it as *BUY*, *SELL*, *HOLD*, or *NONE*. The model also returns a directional score on $[-1, +1]$ that captures the strength of the recommendation (see Table OA.1 in the Online Appendix for the prompt design). This measure is designed to capture the type of actionable guidance that prior work has shown to be price-relevant in the context of analyst recommendations and financial media (Barber et al., 2001; Sinha et al., 2019; Womack, 1996). Among podcast-exposure observations, 46.7% contain at least one buy signal, 16.0% contain a sell signal, and 13.2% contain a hold signal; 34.0% contain no explicit recommendation. The roughly 3:1 buy-to-sell ratio is consistent with the optimism bias documented for other information intermediaries (Barber et al., 2001).

Finally, we measure conviction, i.e. how strongly the speaker appears to stand behind a given view. Our primary measure is a continuous LLM conviction score on $[0, 1]$, obtained by instructing the model to assess whether the speaker expresses high confidence or hedges and qualifies the discussion (see Table OA.1 in the Online Appendix for the prompt design). To verify robustness, we construct several alternatives. First, we compute a TF-IDF conviction score that measures how closely a snippet's wording resembles confident versus hedged reference language, using the same prototype approach as for sentiment and actionability. Second, we count modal strong words (e.g., "will," "always") and uncertainty words (e.g., "may," "possibly") from the

Loughran and McDonald (2011) dictionary. Third, we compute a hedge-to-boost ratio based on custom word lists that capture hedging expressions relative to intensifiers (see variable definition in Online Appendix Table OA.1). Together, these measures allow us to distinguish *what* is being said about a stock from *how strongly* it is being said.

Because a stock can be discussed in multiple snippets on the same day across different episodes or shows, we aggregate snippet-level measures to the stock-day level. For continuous measures (sentiment, forward-looking language, actionability, and conviction) we compute both the unweighted mean and the listener-score-weighted mean across all snippets for a given stock-day. For recommendation measures, we sum the number of BUY, SELL, and HOLD snippets.

To approximate audience reach of a podcast, we use the Listener Score provided in the Listen Notes dataset. Audience reach is not directly observable, as podcast download and listener counts are proprietary and not publicly disclosed by hosting platforms. The Listen Notes listener score is a composite popularity index on a 0–100 scale that aggregates platform-internal engagement signals (e.g., plays and estimated downloads) with external indicators (e.g., online mentions and reviews), conceptually similar to established audience metrics such as Nielsen ratings. Because our sample is restricted to established, equity-focused shows, the observed scores span a relatively narrow range of 41 to 67 (median 54; see Figure A.1 in the Appendix), reflecting that all sample podcasts have substantial but varying audience reach.

Figure 1 provides a time-series overview of podcast sentiment and coverage alongside market returns and realized volatility over the sample period.[6] Panel A shows that podcast sentiment co-moves with market returns at the weekly frequency ($\rho = 0.36$), consistent with hosts responding to and interpreting recent market movements in their discussions. Sentiment drops sharply during market downturns and recovers alongside prices, indicating that the LLM-based measure captures meaningful variation in the tone of stock discussions. Panel B displays realized market volatility, which spikes during stress episodes but is only weakly correlated with sentiment ($\rho = -0.02$). Together, these patterns suggest that while podcast hosts adjust the tone of their discussions in response to market conditions, the overall sentiment of podcast coverage is not merely a byproduct of market-wide events. This distinction motivates our baseline regression analyses, which examine whether podcast coverage has an incremental effect on trading behavior.

[6]To contextualize the data, we map the series against five salient market events: the Volmageddon episode (January–February 2018), the Q4 2018 sell-off, the U.S.–China trade-war escalation (August 2019), the COVID-19 crash (February–March 2020), and the meme-stock episode (January 2021).

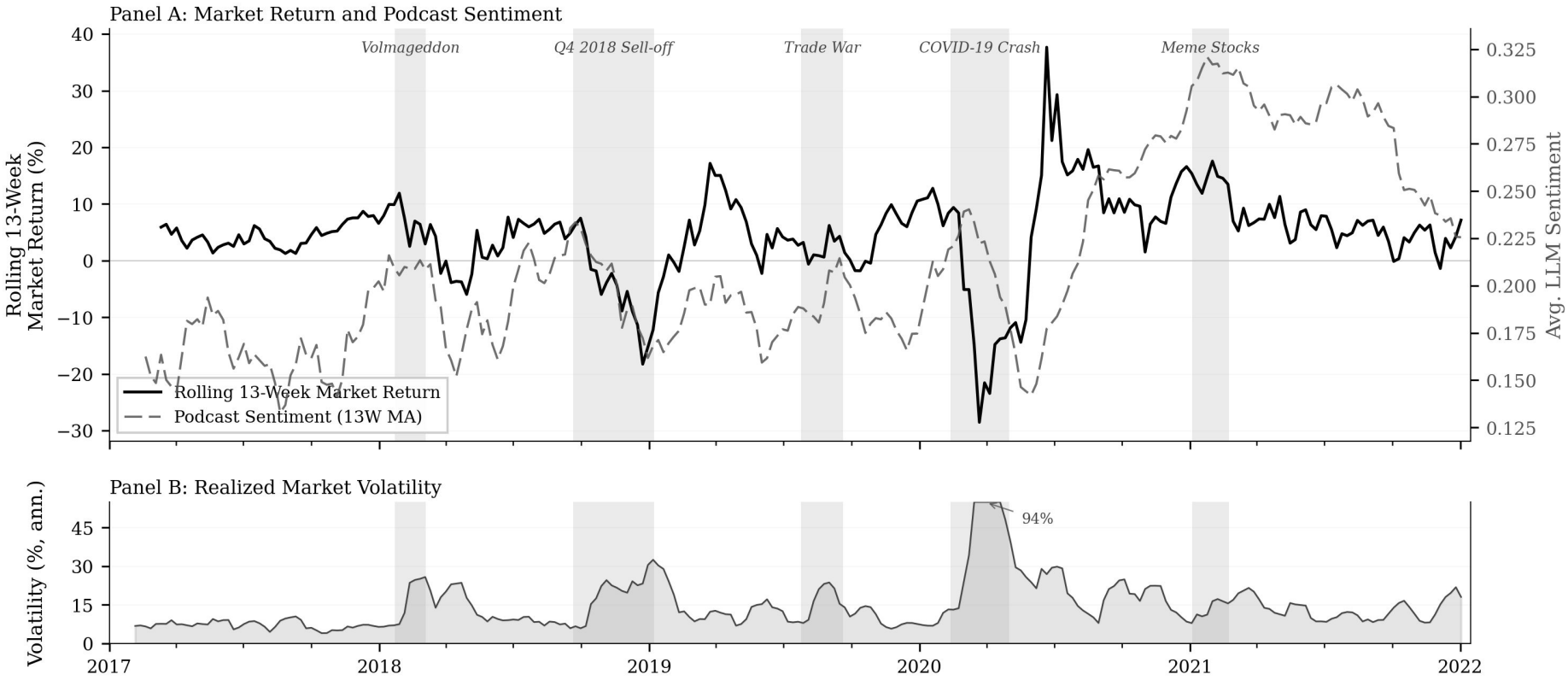


Figure 1: Podcast Sentiment and Coverage Over Time

Notes. This figure plots weekly podcast sentiment alongside market returns and realized volatility from January 2017 to December 2021. Panel A overlays the rolling 13-week compound market return (solid line, left axis) with the cross-sectional average LLM sentiment score smoothed by a 13-week moving average (dashed line, right axis). Panel B displays annualized realized market volatility, computed as the 21-trading-day rolling standard deviation of daily value-weighted market returns ($\times\sqrt{252}$). The volatility axis is capped at 55%; the COVID-19 peak reaches 94%. Gray bands mark five salient market events: Volmageddon (Jan–Feb 2018), the Q4 2018 sell-off (Oct–Dec 2018), the U.S.–China trade-war escalation (Aug 2019), the COVID-19 crash (Feb–Mar 2020), and the meme-stock episode (Jan 2021). Weekly pairwise correlations ($N = 252$): $\rho(\text{Return}, \text{Sentiment}) = 0.36$, $\rho(\text{Return}, \text{Volatility}) = -0.55$, $\rho(\text{Sentiment}, \text{Volatility}) = -0.02$. Market returns are based on the Fama–French value-weighted market factor plus the risk-free rate.

2.3 Market and Firm Data

We merge the podcast panel with several standard financial databases. From CRSP, we obtain daily returns, trading volume, shares outstanding, and market capitalization for all common stocks (share codes 10 and 11) listed on NYSE, AMEX, and NASDAQ. We use this data to construct turnover, defined as daily volume divided by shares outstanding, absolute returns, and Parkinson volatility. From Compustat, we obtain annual book-to-market ratios, leverage, and return on assets, matched through the CRSP–Compustat linking tables. From I/B/E/S, we measure analyst following as the log of one plus the number of analysts issuing EPS forecasts, and we construct earnings surprise measures from actual and consensus forecast data. From Robintrack,[7] we obtain daily ticker-level Robinhood user counts from May 2018 through August 2020 and construct $\Delta \log$ holders, which captures proportional net buying, and the absolute change in holder counts, reflecting trading intensity among retail investors. These data have become a standard source for measuring retail investor activity at the stock-day level, as Robinhood's large and predominantly retail user base provides a high-frequency window into attention-driven trading (Barber et al., 2022; Eaton et al., 2022; Welch, 2022). From SDC Platinum, we obtain M&A announcement dates, deal values, and announcement returns for the identification tests around corporate events. Finally, we use the daily Fama–French five factors plus momentum from Ken French's data library for the portfolio analysis.

2.4 Sample

We construct a stock-day panel by merging CRSP daily data with the stock-day podcast aggregates. We classify a stock-day as a "podcast exposure day" if the stock receives at least one snippet mention on that calendar date. Continuous exposure measures including log listener-weighted mentions, mean sentiment, and related content measures are set to zero on non-exposure days. We winsorize all continuous variables at the 1st and 99th percentiles (see Table OA.1 in the Online Appendix for definitions and construction details of all variables). Lagged controls, including return, turnover, and absolute return, are constructed as one-day lags within each stock.

The final panel comprises 3,527,519 stock-days across 2,850 tickers and spans January 2016 through June 2022. Of these observations, 44,905 (1.27%) are podcast exposure days. The

[7] robintrack.net

Robintrack subsample contains 1,153,389 stock-days across 2,252 tickers, including 19,169 podcast exposure days (1.66%). Table 1 reports summary statistics for the variables used in the analysis.

Table 1: Summary Statistics

Variable	N	Mean	SD	Median	P5	P95
Panel A: Dependent Variables Market						
Turnover	3,527,519	0.0113	0.0157	0.0067	0.0008	0.0362
|Return|	3,527,519	0.0209	0.0239	0.0132	0.0009	0.0689
Daily Return	3,527,519	0.0003	0.0300	0.0000	−0.0486	0.0489
Log Market Cap	3,527,519	21.239	2.159	21.327	17.479	24.699
Panel B: Dependent Variables Retail (Robintrack)						
Δ Log Holders	1,151,282	0.003	0.020	0.000	−0.022	0.034
|Δ Holders|	1,151,282	21.333	75.192	3.000	0.000	88.000
Panel C: Treatment						
Podcast Exposure (0/1)	3,527,519	0.013	0.112	0.000	0.000	0.000
Panel D: Podcast Content (podcast exposure days only)						
LLM Sentiment	44,905	0.257	0.554	0.500	−1.000	1.000
FinBERT Sentiment	44,905	−0.016	0.373	0.030	−0.836	0.627
Forward-Looking (LLM)	44,905	0.338	0.354	0.250	0.000	1.000
Actionability	44,905	0.538	0.024	0.536	0.507	0.575
Recommendation Score	44,905	0.224	0.420	0.000	−0.500	0.800
Buy Signal (0/1)	44,905	0.467	0.499	0.000	0.000	1.000
Sell Signal (0/1)	44,905	0.160	0.367	0.000	0.000	1.000
Hold Signal (0/1)	44,905	0.132	0.339	0.000	0.000	1.000
Conviction (LLM)	44,905	0.691	0.225	0.800	0.250	1.000
Listener Score	44,905	54.579	7.186	54.000	43.000	66.000
LS-Weighted Mentions	44,905	95.780	90.439	60.000	44.000	255.000

Notes: This table reports summary statistics for the master analysis panel (January 2016–June 2022). Panel A reports market-level outcomes from CRSP. Turnover is daily volume divided by shares outstanding. Panel B reports retail outcomes from Robintrack (May 2018–August 2020). Panel C reports the binary treatment indicator. Panel D reports podcast content variables, measured only on podcast exposure days ($N = 44{,}905$). LLM Sentiment is the stock-day mean LLM-based sentiment score (−1 to +1). Forward-looking, actionability, recommendation, and conviction are LLM-derived content dimensions. Listener score measures podcast audience reach. All continuous variables are winsorized at the 1st and 99th percentiles. See Table OA.1 in the Online Appendix for detailed variable definitions.

Listener scores range from 41 to 67, with a median of 54 and a mean of 55 (see Figure A.1 in the appendix for the full distribution).

3 Trading Response to Podcast Coverage

3.1 Baseline Regressions

We begin by examining whether podcast coverage predicts increased investor activity on the day of the mention. We consider four dependent variables that capture complementary dimensions of the response.

On the market side, we use turnover, defined as daily share volume divided by shares outstanding, and absolute returns. Turnover is a standard measure of trading activity in the market microstructure literature (Chordia et al., 2001; Lo & Wang, 2000). Absolute returns capture the magnitude of price movements, independent of direction, and are commonly used to measure the extent of investor reaction (Ben-Rephael et al., 2017; Da et al., 2011). On the retail side, we use Robintrack data on the daily number of Robinhood users holding each stock. From these data, we construct $\Delta \log$ holders, which captures proportional net buying or selling by retail investors, and the absolute change in holder counts, which captures the intensity of retail trading irrespective of direction (Barber et al., 2022; Eaton et al., 2022; Welch, 2022). This allows us to assess whether the response to podcast coverage is visible in both broad market activity and retail trading.

We estimate panel regressions of the form:

$$y_{i,t} = \beta \cdot PodcastExposure_{i,t} + \gamma' X_{i,t-1} + \alpha_i + \boldsymbol{\delta}_{\text{cal}(t)} + \varepsilon_{i,t} \tag{1}$$

where $y_{i,t}$ is the outcome for stock i on day t, α_i denotes stock fixed effects, $\boldsymbol{\delta}_{\text{cal}(t)}$ is a vector of calendar fixed effects (day-of-week, week-of-month, month-of-year, and year), and $X_{i,t-1}$ is the set of lagged controls. Standard errors are clustered at the stock level. The calendar fixed effects absorb systematic trading patterns at common calendar frequencies. The control vector includes lagged return, lagged turnover, lagged absolute return, log market capitalization, book-to-market ratio, leverage, return on assets, and log analyst following (see Table OA.1 in the Online Appendix for detailed variable definitions).

Table 2 reports the baseline results. Podcast coverage is associated with higher investor activity across all four outcomes. Turnover is 0.34 percentage points higher on podcast exposure days ($t = 15.81$), which corresponds to about 30% of its unconditional mean of 1.13%. Absolute returns increase by 0.53 percentage points ($t = 23.08$). On the retail side, $\Delta \log$ holders is 0.52% ($t = 20.05$) higher, and the absolute change in holder counts increases by 45 ($t = 17.63$). These patterns indicate that podcast coverage coincides with both broader market activity and retail trading.

To examine timing more closely, we estimate event-time regressions that trace the podcast-

Table 2: Podcast Coverage and Investor Activity

	(1) Turnover	(2) |Ret|	(3) Δ Log Hold.	(4) |Δ Hold.|
Podcast Exposure	0.0034***	0.0053***	0.0052***	45.32***
	(0.0002)	(0.0002)	(0.0003)	(2.57)
Controls	Included			
Fixed Effects	Stock + Calendar (DOW, WoM, Month, Year)			
Clustering	Stock			
Sample	Full Panel		Robintrack Subsample	
N	3,524,669	3,524,669	1,151,131	1,151,131
R^2 (within)	0.472	0.042	0.005	0.111

Notes: This table reports panel regressions of investor activity on the podcast-exposure indicator. Columns (1)–(2) use the full CRSP panel; columns (3)–(4) restrict to the Robintrack subsample (May 2018–August 2020). Calendar fixed effects absorb day-of-week, week-of-month, month-of-year, and year effects. Controls include lagged return, lagged turnover, lagged absolute return, log market cap, book-to-market, leverage, ROA, and log analyst following. All continuous variables are winsorized at the 1st and 99th percentiles. Standard errors clustered at the stock level are in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels. See Table OA.1 in the Online Appendix for detailed variable definitions.

exposure coefficient across leads ($t-5$ to $t-1$) and lags ($t+1$ to $t+5$):

$$y_{i,t} = \sum_{k=-5}^{+5} \beta_k \cdot \mathbf{1}[\text{Podcast}_{i,t+k}] + \gamma' X_{i,t-1} + \alpha_i + \boldsymbol{\delta}_{\text{cal}(t)} + \varepsilon_{i,t} \tag{2}$$

All specifications use the full control set. Figure 2 plots the coefficient paths for all four outcomes. For turnover and absolute returns, the lead coefficients are close to zero and partly slightly negative until the podcast exposure day. This pattern is plausible if podcast mentions cluster around scheduled information events such as earnings announcements: prior work shows that trading volume tends to fall before major scheduled news events as information asymmetry rises (Chae, 2005; Levi & Zhang, 2015; So & Wang, 2014). For the retail measures, the coefficients begin to rise modestly before $t = 0$, especially near $t-1$. This is consistent with pre-event speculation by individual investors. Robinhood users have been shown to engage in attention-induced trading, and retail order imbalance increases ahead of major news events (Barber et al., 2001; Liu et al., 2020). This pattern suggests that podcast coverage often occurs around periods of increasing investor attention rather than arriving in isolation.

At the same time, all four outcomes display a clear discrete increase at $t = 0$. The coefficients then decline over the following days. For the market outcomes, this decline is fairly rapid. For the retail measures, it is more persistent, which is consistent with podcast content diffusing over

several days as listeners process and act on it with some delay. Overall, the event-time evidence shows that podcast coverage coincides closely with elevated trading, while also indicating that podcasts tend to cluster around periods of already elevated attention.

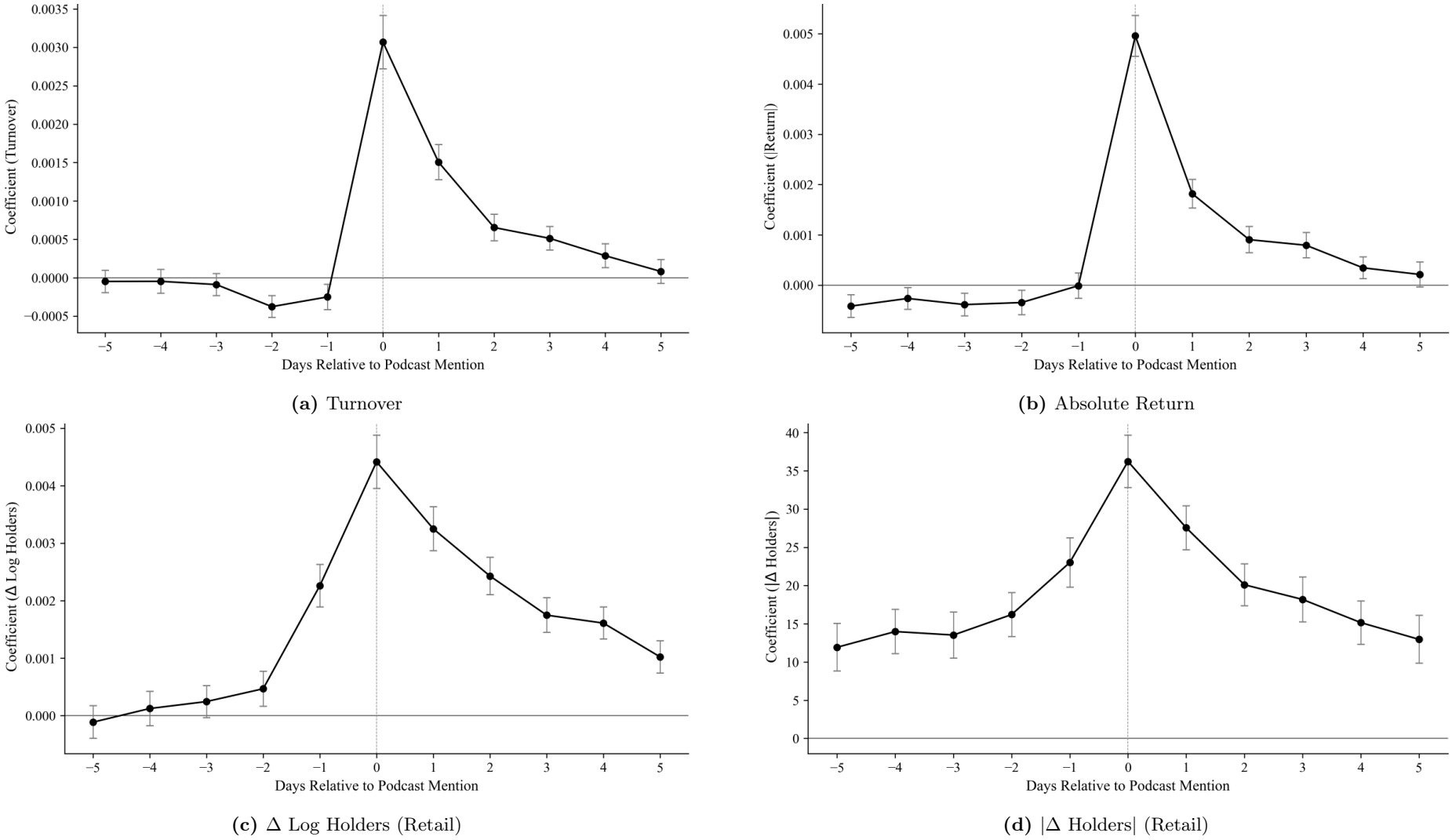
3.2 Cadence of Episode Releases

A potential concern is that podcast episodes may be timed to coincide with newsworthy events, so that the estimated podcast effect partly captures the market response to the underlying news rather than to the podcast itself. We exploit the regular release cadence of podcast shows to address this concern. Many investing podcasts follow a fixed weekly schedule, releasing new episodes on the same weekday regardless of what is happening in the market. For these shows, the timing of episode availability is determined by the production schedule, not by stock-specific developments.

For each show, we compute the empirical distribution of episode releases across weekdays and identify the show's peak release day. We then classify each podcast mention as *on-schedule* if the episode aired on the show's regular publication day, or *off-schedule* if the episode aired on a different day. On-schedule episodes are more likely to reflect supply-driven availability, where the show was going to air that day regardless of what happened to any particular stock. Off-schedule episodes, by contrast, may reflect demand-driven production decisions, where hosts publish an extra or rescheduled episode in response to market events.

Table A.3 in the Appendix decomposes the baseline podcast-exposure effect by release cadence. Both types of mentions enter the same regression, with non-exposure days as the reference category. The effects of on-schedule mentions are smaller than those for off-schedule mentions, consistent with off-schedule episodes containing more event-driven and therefore more market-moving content. However, the on-schedule effect remains large and significant.

These results suggest that even when podcast exposure is plausibly supply-driven, arising from a show's pre-determined release schedule rather than from a news event, the trading response remains economically and statistically significant.



(a) Turnover

(b) Absolute Return

(c) Δ Log Holders (Retail)

(d) |Δ Holders| (Retail)

Figure 2: Event-Time Coefficients Around Podcast Exposure Events

Notes: Coefficient estimates from Equation (2) for four dependent variables. Each panel plots lead/lag coefficients ($t-5$ to $t+5$) of the podcast-exposure indicator. Filled circles are point estimates connected by lines; whiskers show 95% confidence intervals. The dashed vertical line marks $t=0$ (podcast exposure day). Controls include lagged return, lagged turnover, lagged absolute return, log market cap, BTM, leverage, ROA, and log analyst following. Stock and calendar fixed effects (DOW, WoM, month, year) are absorbed; standard errors are clustered by stock. Panels (c) and (d) use the Robintrack subsample (May 2018–August 2020).

3.3 Heterogeneity in Investor Responses

The baseline results establish a strong relation between podcast coverage and investor activity. However, this relation may reflect confounding by underlying news rather than an effect of podcast coverage itself. For example, corporate announcements may act as confounders if they simultaneously affect both trading and podcast discussion, leading the estimated podcast-exposure effect to capture a shared response rather than an incremental podcast effect. The tests in this section examine whether the observed patterns are consistent with podcasts exerting an incremental effect on investor behavior.

Under a pure common-news confounder, the underlying event should account for both the podcast mention and the trading response. In that case, differences in podcast reach or in how the stock is framed should not be influential for the size of the response. If podcasts exert an independent influence on investor behavior, the response should be stronger when coverage reaches more listeners, contains more forward-looking discussion, provides more actionable language, or includes explicit buy or sell recommendations. These signal characteristics are difficult to attribute to a common-news confounder, because they reflect how podcast content is framed and transmitted rather than the occurrence of the underlying news itself. Instead, under an incremental podcast effect these characteristics should matter because they govern how widely and how effectively the information is transmitted to investors. They therefore provide a useful setting in which to examine whether podcasts have an incremental effect on investor behavior beyond any shared response to common news. We interact the podcast-exposure indicator with four content moderators:

$$y_{i,t} = \beta_1 \cdot Podcast_{i,t} + \beta_2 \cdot Podcast_{i,t} \times Moderator_{i,t} + \gamma' X_{i,t-1} + \alpha_i + \boldsymbol{\delta}_{\mathrm{cal}(t)} + \varepsilon_{i,t} \qquad (3)$$

where $Moderator_{i,t}$ denotes one of four podcast characteristics. *Listener score* captures audience reach using the proprietary Listen Notes popularity index. *Forward-looking language* measures the extent to which the discussion focuses on future developments rather than past events. *Actionability* captures how closely the language resembles an investment recommendation. *Buy/Sell recommendation* is an indicator equal to one if the podcast-exposure contains at least one explicit buy or sell signal. We estimate each interaction separately. Continuous moderators are standardized on podcast exposure days to have mean zero and unit variance.

All other variables and controls are defined as in Equation (1); see Table OA.1 in the Online Appendix for detailed definitions of all moderator variables.

Table 3: Heterogeneity in Market Responses to Podcast Coverage

	Listener Score (1)	Forward-Looking (2)	Action-ability (3)	Buy/Sell Signal (4)
Panel A: Turnover				
Podcast Exposure (β_1)	0.00349***	0.00337***	0.00338***	0.00226***
	(0.00024)	(0.00021)	(0.00022)	(0.00017)
Podcast × Moderator (β_2)	0.00160***	0.00028***	0.00131***	0.00193***
	(0.00022)	(0.00007)	(0.00013)	(0.00021)
	[7.3]	[3.8]	[10.4]	[9.1]
Panel B: Absolute Return				
Podcast Exposure (β_1)	0.00552***	0.00532***	0.00534***	0.00316***
	(0.00028)	(0.00023)	(0.00024)	(0.00023)
Podcast × Moderator (β_2)	0.00258***	0.00030**	0.00223***	0.00375***
	(0.00027)	(0.00012)	(0.00021)	(0.00028)
	[9.4]	[2.5]	[10.7]	[13.6]
Controls	Included			
Fixed Effects	Stock + Calendar (DOW, WoM, Month, Year)			
Clustering	Stock			
N	3,524,669			

Notes: Each column reports a separate regression interacting the podcast-exposure indicator with the listed moderator. Columns (1)–(3): moderators are standardized to zero mean and unit variance on podcast exposure days. Column (4): Buy/Sell Signal is a dummy equal to one if the podcast mention contains at least one explicit buy or sell recommendation (58.3% of podcast exposure days). t-statistics in brackets. ***, **, and * denote significance at the 1%, 5%, and 10% levels. See Table OA.1 in the Online Appendix for detailed variable definitions.

Tables 3 and 4 report the results. Each column reports a separate regression in which the podcast-exposure indicator is interacted with one moderator. Columns (1) through (3) use continuous moderators standardized to have mean zero and unit variance on podcast exposure days. Column (4) uses an indicator for the presence of explicit buy or sell recommendations.

The interaction coefficients are positive and statistically significant across most specifications. Stocks discussed by podcasts with larger audiences exhibit stronger responses in both market-wide and retail outcomes. The same pattern appears when podcast coverage is more forward-looking, more actionable, or contains an explicit directional recommendation. Across specifications, the coefficient on the podcast-exposure indicator remains positive, while the interaction terms show that the response is stronger when the podcast is more likely to deliver interpretable and decision-relevant information.

These patterns are difficult to reconcile with a common-news explanation alone. If an underlying event fully drove both podcast mentions and investor activity, differences in audience

Table 4: Heterogeneity in Retail Trading Responses to Podcast Coverage

	Listener Score (1)	Forward-Looking (2)	Action-ability (3)	Buy/Sell Signal (4)
Panel A: Δ Log Holders				
Podcast Exposure (β_1)	0.00551***	0.00520***	0.00525***	0.00395***
	(0.00030)	(0.00026)	(0.00027)	(0.00028)
Podcast × Moderator (β_2)	0.00313***	0.00039**	0.00106***	0.00220***
	(0.00040)	(0.00016)	(0.00021)	(0.00038)
	[7.8]	[2.4]	[5.1]	[5.8]
Panel B: |Δ Holders|				
Podcast Exposure (β_1)	48.423***	45.523***	45.627***	33.457***
	(2.515)	(2.548)	(2.563)	(2.840)
Podcast × Moderator (β_2)	30.581***	5.646***	5.341***	21.083***
	(2.468)	(1.102)	(1.217)	(2.728)
	[12.4]	[5.1]	[4.4]	[7.7]
Controls	Included			
Fixed Effects	Stock + Calendar (DOW, WoM, Month, Year)			
Clustering	Stock			
N	1,151,131			

Notes: Each column reports a separate regression interacting the podcast-exposure indicator with the listed moderator, using the Robintrack subsample. Specification details are as in Table 3. ***, **, and * denote significance at the 1%, 5%, and 10% levels. See Table OA.1 in the Online Appendix for detailed variable definitions.

reach or in how the podcast frames the stock should not matter for the size of the response. Instead, the evidence suggests that podcast coverage contributes incrementally to investor trading beyond any shared response to common news.

3.4 Incremental Diffusion Around Corporate News Events

The evidence so far suggests that podcast coverage is linked to investor trading in a way that goes beyond a pure common-news confounder. At the same time, the pre-trends in the event-time patterns in Section 3.1 indicate that podcast mentions seem to co-occur around periods of heightened attention. These patterns are consistent with the possibility that podcasts amplify investor responses to firm-specific news once that news is already in circulation. Such a role would be consistent with an information-diffusion channel, in which podcasts widen the reach of firm-specific news, reduce processing costs, and help investors interpret information in a more actionable form. We examine this mechanism more directly by studying podcast coverage around predetermined corporate news events, specifically earnings announcements and M&A announcements.

These settings are useful because the underlying news event is observed directly and its

timing is predetermined. This allows us to compare the market response to the same type of public news with and without podcast coverage. If podcasts merely echo public news without affecting its transmission to investors, then podcast coverage around such events should exhibit no incremental effect on investor trading. If, by contrast, podcasts help diffuse, interpret, or contextualize public news, the trading response should be stronger when the event is accompanied by podcast coverage.

We test this idea by estimating the following specification:

$$y_{i,t} = \beta_1 \cdot EA_{i,t} + \beta_2 \cdot Podcast_{i,t} + \beta_3 \cdot (EA_{i,t} \times Podcast_{i,t}) + \gamma' X_{i,t-1} + \alpha_i + \boldsymbol{\delta}_{\text{cal}(t)} + \varepsilon_{i,t} \quad (4)$$

where $EA_{i,t}$ is an indicator for an earnings announcement window (± 1 trading day around the I/B/E/S-reported announcement date). The coefficient β_1 captures the average response to the earnings event itself. The coefficient β_2 captures the podcast-exposure effect outside earnings windows. The interaction term β_3 is the coefficient of interest. It measures whether investor response is stronger when a stock is discussed in a podcast on or around an earnings announcement. Under a pure common-news explanation, β_3 should be null once the earnings event is controlled for. Under an incremental diffusion interpretation, β_3 should be positive because podcasts broaden the reach of earnings news or make it easier for investors to process and act on it.

Table 5 shows that podcast coverage is associated with a stronger investor response around earnings announcements. The interaction coefficient, β_3, is positive and statistically significant for all four outcomes, even after controlling for the earnings surprise itself and its absolute value. This pattern indicates that the response on earnings-announcement days is stronger when the stock is also discussed in a podcast. Figure 3 provides an exemplary visualization of this effect for the turnover as the dependent variable. While turnover increases around earnings announcements even in the absence of podcast coverage, the increase is substantially larger when the same event is accompanied by podcast discussion.

The economic magnitudes are substantial. The combined effect on earnings-announcement days with podcast coverage, $\beta_1 + \beta_2 + \beta_3$, is considerably larger than the earnings-announcement effect without podcast coverage, β_1. For turnover, an earnings announcement accompanied by podcast coverage is associated with 141% more trading (even after controlling for stock, firm, calendar, and earnings characteristics) than an earnings announcement without podcast

Table 5: Incremental Podcast Effects Around Earnings Announcements

	Turnover (1)	|Return| (2)	Δ Log Hold. (3)	|Δ Hold.| (4)
EA Window (β_1)	0.00597***	0.01248***	0.01150***	16.536***
	(0.00012)	(0.00014)	(0.00026)	(0.732)
	[51.1]	[86.1]	[44.2]	[22.6]
Podcast Exposure (β_2)	0.00223***	0.00362***	0.00402***	33.338***
	(0.00018)	(0.00020)	(0.00025)	(2.756)
	[12.5]	[18.2]	[16.3]	[12.1]
EA × Podcast (β_3)	0.00616***	0.00612***	0.00191**	90.563***
	(0.00066)	(0.00080)	(0.00091)	(6.248)
	[9.3]	[7.7]	[2.1]	[14.5]
Earnings Surprise	-0.00003	-0.00007**	0.00006	-0.202
	(0.00003)	(0.00003)	(0.00004)	(0.126)
|Earnings Surprise|	-0.00001**	-0.00002***	-0.00005***	-0.108***
	(0.00000)	(0.00001)	(0.00001)	(0.039)
Implied effects on earnings days:				
EA without podcast (β_1)	0.00597	0.01248	0.01150	16.536
EA with podcast ($\beta_1 + \beta_2 + \beta_3$)	0.01436	0.02222	0.01743	140.437
Difference ($\beta_2 + \beta_3$)	+0.00839	+0.00974	+0.00593	+123.901
Wald F-statistic	134.6***	139.2***	41.0***	432.6***
Δ%	+141%	+78%	+52%	+749%
Controls	Baseline set + Earnings Surprise			
Fixed Effects	Stock + Calendar (DOW, WoM, Month, Year)			
Clustering	Stock			

Notes: Each column reports a single regression of the form in Equation (4), augmented with earnings surprise controls. EA Window equals one if the stock has an I/B/E/S-reported quarterly earnings announcement within ±1 trading day. Earnings Surprise is actual EPS minus the consensus mean analyst forecast (I/B/E/S); |Earnings Surprise| is its absolute value. Both are winsorized at the 1st and 99th percentiles and set to zero on non-EA days. Of 44,905 podcast exposure stock-days, 4,877 (10.9%) fall within an EA window. The lower panel reports implied effects: "EA without podcast" is $\hat{\beta}_1$; "EA with podcast" is $\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$. The Wald F-test tests $H_0 : \beta_2 + \beta_3 = 0$. Δ% is the percentage increase from EA-only to EA-with-podcast. Columns (3)–(4) use the Robintrack subsample. t-statistics in brackets. ***, **, and * denote significance at the 1%, 5%, and 10% levels. See Table OA.1 in the Online Appendix for detailed variable definitions.

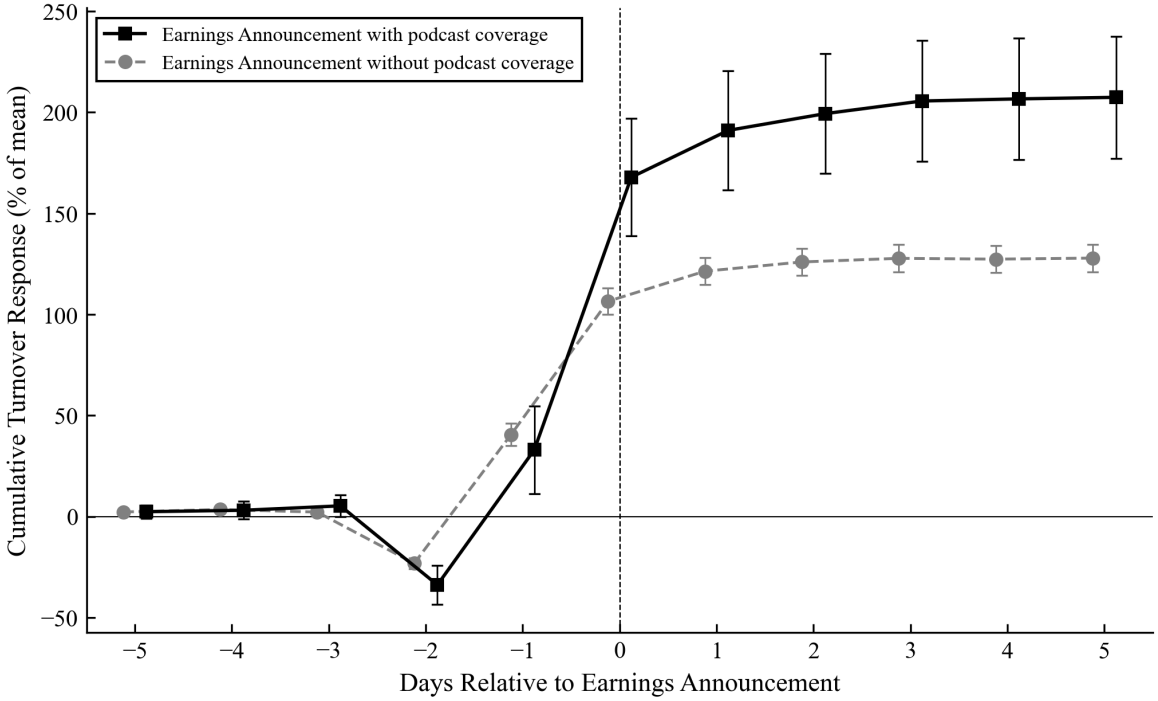


Figure 3: Turnover Response Around Earnings Announcements With and Without Podcast Coverage

Notes: This figure plots the cumulative turnover response, expressed as a percentage of mean turnover, from five trading days before to five trading days after the earnings announcement date. The solid line with square markers shows earnings announcements with podcast coverage; the dashed line with circular markers shows earnings announcements without podcast coverage. The vertical dashed line marks day 0, the earnings announcement date. Turnover rises around earnings announcements in both cases, but the increase is larger and more persistent when the event is accompanied by podcast coverage. Regression specification follows equation 4 with varying t.Table 5 depicts results for all depended variables

coverage ($F = 134.6$, $p < 0.001$). For absolute retail activity, the corresponding difference is 749% ($F = 432.6$, $p < 0.001$). Similar differences arise for the other outcomes.

We extend the news-event framework to mergers and acquisitions (M&A), using deal announcements from SDC Platinum. M&A events provide a second source of exogenous, predetermined corporate events whose timing cannot be caused by podcast coverage. We estimate the same interaction specification as Equation (4), replacing the earnings indicator with an M&A announcement window (±1 trading day). To control for deal characteristics, we add log deal value (and absolute 1-day announcement return) as a proxy for deal size and attention (positivity of market reaction) as independent variables.

Table A.4 in the Appendix shows that the podcast amplification effect also extends to M&A events. The interaction coefficient β_3 is positive and significant for all four outcomes, even after controlling for deal size and announcement returns. M&A announcement days accompanied by podcast coverage are associated with substantially stronger trading responses than M&A days without podcast coverage. For turnover, the Wald test rejects equality at $F = 47.6$ ($p < 0.001$).

Our earlier findings show that podcast coverage closely coincides with elevated trading activity and that this relation is not driven solely by common-news confounding. The results here point to the channel through which this incremental effect appears to arise. Conditional on the same type of predetermined corporate news event, investor response is stronger when podcast coverage is present. This pattern is consistent with podcasts amplifying the diffusion of corporate news by broadening its reach, lowering processing costs, or helping investors interpret the information in a more actionable form.

3.5 Retail-Skewed Trading Responses

If podcasts function as an information-diffusion channel, their effects should be most visible where information acquisition and processing costs are highest. Retail investors have more limited access to specialized research infrastructure and real-time news analytics than institutional investors and may therefore be more likely to rely on accessible media such as podcasts (Boehmer et al., 2021; Eaton et al., 2022). This suggests that trading responses to podcast coverage should be relatively more pronounced in retail activity than in broader market activity.

Because our retail and market outcomes are measured in different units (Δ log holders vs. turnover), direct coefficient comparisons across separate regressions are not informative.[8] To

[8]More generally, formal comparisons of regression coefficients across models require care when scale differences

obtain an indicative comparison, we construct a joint classification scheme. For each stock-day in the Robintrack subsample, we compute the within-day percentile rank of turnover (market activity) and $|\Delta \text{ holders}|$ (retail activity). A stock-day is classified as "high" on a given dimension if it falls above the daily cross-sectional median, yielding four quadrants: *Retail-Dominant* (high retail, low market), *Both Active* (high retail, high market), *Market-Dominant* (low retail, high market), and *Both Quiet* (low retail, low market). The two asymmetric quadrants (Retail-Dominant and Market-Dominant) are of primary interest: if podcasts disproportionately reach retail investors, podcast exposure days should be more strongly associated with the Retail-Dominant than the Market-Dominant quadrant.

Table 6: Podcast Reactions Among Retail Investors versus the Broader Market

Day Type	(1) Retail-Dominant	(2) Both Active	(3) Market-Dominant	(4) Both Quiet
Share of stock-days	17.5%	31.8%	18.3%	32.5%
Podcast Exposure	0.10222^{***}	0.10478^{***}	-0.16129^{***}	-0.04571^{***}
	(0.01684)	(0.01536)	(0.00667)	(0.01327)
	[6.1]	[6.8]	[−24.2]	[−3.4]
Retail-Dominant vs. Market-Dominant:				
$\beta_{(1)} - \beta_{(3)}$		+0.264 ($t = 13.78$, $p < 0.001$)		
Controls	Lagged return, lagged turnover, log market cap			
Fixed Effects	Calendar (DOW, WoM, Month, Year)			
Clustering	Stock			
N	1,153,149			

Notes: This table reports linear probability model (LPM) estimates of podcast exposure on the likelihood that a stock-day falls into each of four trading activity quadrants. Quadrants are defined by daily cross-sectional median ranks (P_{50}) of market turnover and absolute change in Robinhood holders. Standard errors clustered by stock are in parentheses; t-statistics in brackets. This specification uses calendar fixed effects without stock fixed effects, as the classification is based on cross-sectional ranks within each trading day. Robintrack subsample (May 2018–August 2020). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 6 reports linear probability models for each quadrant indicator. Podcast coverage is associated with a higher probability of falling into the Retail-Dominant quadrant by 10.2 pp ($t = 6.1$) and the Both Active quadrant by 10.5 pp ($t = 6.8$), while it reduces the probability of Market-Dominant days by 16.1 pp ($t = -24.2$). The difference between the Retail-Dominant and Market-Dominant coefficients is large and highly significant ($\beta_{(1)} - \beta_{(3)} = +0.264$, $t = 13.78$, $p < 0.001$). On non-exposure days, the four quadrants are roughly symmetrically distributed; on podcast exposure days, the distribution shifts sharply toward the two high-retail quadrants and away from Market-Dominant days.

affect their interpretation (Clogg et al., 1995; Paternoster et al., 1998).

We note that this classification-based approach provides suggestive rather than conclusive evidence, as it compares relative rankings rather than effect magnitudes. With this caveat, the results indicate that podcast coverage is associated with disproportionately strong retail trading responses, while market-only reactions become substantially less likely.

3.6 Sentiment Decomposition and Negativity Bias

Podcasts seem to transmit stock-related information to (especially retail) investors, who then respond to such coverage. Thus, established behavioral regularities in how investors react to information should also appear in our setting.

A robust finding in behavioral finance is that investors react more strongly to negative than to positive information. Prospect theory (Kahneman & Tversky, 1979) predicts that losses loom larger than equivalent gains, making negative information more salient and action-inducing. The broader psychological literature documents a pervasive negativity bias: negative events carry greater weight in evaluations, attract more attention, and are processed more thoroughly than positive events of comparable magnitude (Baumeister et al., 2001). Related evidence in finance shows that negative media tone and negative information shocks are associated with stronger market reactions and more persistent adjustment than comparable positive information (Jiang & Zhu, 2017; Tetlock, 2007; Tetlock et al., 2008). With podcasts functioning as a channel of information diffusion, we should observe the same asymmetry here: negative podcast exposure days should trigger a stronger investor response than positive podcast exposure days.

We test this prediction by decomposing the podcast-exposure indicator into positive and negative sentiment days based on the stock-day mean LLM sentiment score (positive > 0.1, negative < -0.1).[9] Both dummies are included in a single regression, and we formally test $H_0 : \beta_{\text{neg}} = \beta_{\text{pos}}$ using Wald tests.

Table 7 shows a clear asymmetry between positive and negative podcast coverage. Both positive and negative podcast exposure days are associated with significant investor responses, indicating that the extensive margin of podcast coverage matters. At the same time, the coefficient on negative podcast exposure days exceeds the coefficient on positive podcast exposure days by 29–67% across the four outcomes, and the Wald tests reject equality in all cases. This pattern is consistent with a negativity bias in how investors respond to podcast content.

[9] The results also hold for FinBERT as an alternative sentiment measure and defining positive (negative) sentiment as values > 0 (< 0).

Table 7: Asymmetric Trading Responses to Positive and Negative Podcast Content

	Turnover (1)	|Return| (2)	Δ Log Hold. (3)	|Δ Hold.| (4)
Positive Day (β_{pos})	0.00314^{***}	0.00516^{***}	0.00450^{***}	41.050^{***}
	(0.00021)	(0.00024)	(0.00028)	(2.547)
	[14.9]	[21.5]	[16.2]	[16.1]
Negative Day (β_{neg})	0.00437^{***}	0.00664^{***}	0.00753^{***}	60.323^{***}
	(0.00031)	(0.00039)	(0.00051)	(4.603)
	[13.9]	[17.0]	[14.9]	[13.1]
$\beta_{\text{neg}} > \beta_{\text{pos}}$: Difference	+39%	+29%	+67%	+47%
Wald F-statistic	26.1^{***}	16.8^{***}	34.6^{***}	18.8^{***}
Controls	Included			
Fixed Effects	Stock + Calendar (DOW, WoM, Month, Year)			
Clustering	Stock			

Notes: Each column reports a single regression replacing the podcast-exposure indicator with two mutually exclusive sentiment dummies based on the stock-day mean LLM sentiment score (positive > 0.1, negative < -0.1; neutral days omitted as baseline). The Wald F-test tests $H_0 : \beta_{\text{neg}} = \beta_{\text{pos}}$. Columns (3)–(4) use the Robintrack subsample. Controls, fixed effects, and clustering are as in Table 2. ***, **, and * denote significance at the 1%, 5%, and 10% levels. See Table OA.1 in the Online Appendix for detailed variable definitions.

3.7 Robustness

3.7.1 Placebo and Falsification Tests

We employ a battery of robustness checks. Three falsification tests indicate that the baseline podcast-exposure effect is unlikely to be a statistical artifact.[10] We conduct each test for all four dependent variables to assess whether the baseline pattern is robust across outcomes.

Second, we randomly reassign podcast exposure days within each stock while preserving the stock-level frequency of podcast coverage and re-estimate the baseline regression 200 times. This test asks whether the results could arise mechanically from certain stocks being mentioned more often than others, independent of the actual timing of podcast coverage. The resulting null distributions are centered near zero, while the actual coefficients lie well outside the placebo distributions for all four outcomes (see Figure A.2 in the appendix).

Third, we assign the same number of podcast exposure days to random calendar dates, again preserving stock-level frequency, across 200 independent draws. This test asks whether the baseline results could be generated by random timing patterns or common calendar structure rather than by the observed timing of podcast mentions. Here too, the null distributions are centered near zero, and the actual estimates lie far outside the placebo distributions for all four

[10] As discussed in the event-time analysis, the lead coefficients at $t-5$ through $t-3$ are economically small relative to the coefficient at $t = 0$ across all four outcomes (Figure 2), which is inconsistent with strong anticipation effects.

outcomes (see Figure A.3 in the appendix).

3.7.2 Alternative Podcast Exposure Measures

The baseline specification uses a binary podcast-exposure indicator as the treatment variable. To assess whether the results depend on this particular operationalization, we re-estimate the baseline regressions using seven alternative continuous measures of podcast exposure. These measures capture different dimensions of podcast activity. Log listener-score-weighted mentions weights each snippet by the audience reach of its source podcast. Log snippet count captures the volume of stock-specific coverage. Log episode count and log show count capture the breadth of coverage across episodes and distinct programs. Absolute LLM sentiment captures the extremity of the podcast signal regardless of direction. The conviction-weighted signal incorporates the confidence of the speaker's language. Log total words captures the depth of discussion. All measures are set to zero on non-exposure days.

Table 8: Alternative Podcast Exposure Measures

Exposure Measure	Turnover t	|Return| t	Δ Log Hold. t	|Δ Hold.| t
Podcast Exposure (0/1)	15.8	23.1	20.1	17.8
Log LS-Wtd Mentions	15.5	22.7	20.0	18.6
Log Snippets	13.6	18.5	17.2	20.1
Log Episodes	13.5	19.3	18.1	20.0
Log Shows	13.5	19.4	18.0	20.1
Abs. Sentiment	15.7	22.2	16.8	17.1
Conviction Signal	15.4	22.9	18.8	17.8
Log Total Words	15.7	22.8	20.2	19.0

Notes: Each cell reports the t-statistic on the exposure measure from the baseline regression (Equation 1) with the full control set, stock and calendar fixed effects, and stock-clustered standard errors. All 32 specifications (8 measures × 4 outcomes) produce positive, significant coefficients at the 1% level. See Table OA.1 in the Online Appendix for detailed variable definitions.

Table 8 reports the results. All 32 specifications, eight exposure measures across four dependent variables, produce positive and significant coefficients at the 1% level. The t-statistics range from 13.5 (log shows predicting turnover) to 23.1 (podcast exposure day predicting absolute returns). The consistency across measures that capture different aspects of podcast exposure show that the effect of our main results is not a mere artifact of any particular operationalization of the variable of interest.

Beyond alternative exposure measures, we verify that the content-based results are robust to the choice of text classification method. The sentiment decomposition results (Section 3.6) are qualitatively unchanged when replacing the LLM sentiment score with FinBERT, the Loughran–

McDonald dictionary, or TF-IDF prototype sentiment (Table A.5 in the Appendix). The forward-looking and conviction results are similarly robust to substituting the primary LLM-based measures with their respective alternatives, namely FinBERT-FLS and dictionary-based proxies for forward-looking language; TF-IDF prototypes and Loughran–McDonald uncertainty counts for conviction (untabulated).

3.7.3 Robustness to Research Design Choices

We further assess robustness by examining the sensitivity of the baseline result to a broad set of reasonable research design choices. Table A.2 summarizes the six dimensions of variation. We vary the exposure measure across all eight alternatives documented in Section 3.7.2, the outcome variable across all four dependent variables, the fixed effects structure, the control set (we use an expanded set in all main analyses but vary to a ore sparse one here), the sample (full panel, large-cap stocks above the median market capitalization, or excluding microcaps below the 20th percentile), and the winsorization level (1% or 0.5%). The full factorial design yields $8 \times 4 \times 2 \times 2 \times 3 \times 2 = 1{,}152$ unique specifications.

Figure A.4 summarizes the resulting distribution of estimates across these specifications. The baseline result remains positive throughout, and the minimum t-statistic ($t = 3.38$) is still statistically significant at conventional levels. All 1,152 specifications produce positive and significant coefficients ($p < 0.05$), with a median t-statistic of 15.28. This indicates that the main finding is not driven by a narrow set of modeling choices. Tightening the winsorization threshold from 1% to 0.5% has little effect on the results, and excluding microcaps strengthens the estimates slightly. Overall, the evidence indicates that the investor-response result is robust to alternative choices of exposure measure, outcome variable, fixed effects, controls, sample restriction, and winsorization.

3.7.4 Coverage Concentration

Podcast coverage is concentrated in a relatively small set of stocks. The 50 most-discussed stocks (2% of tickers) account for 41.5% of all podcast exposure stock-days, and the top 100 (4%) account for 56.1%. The median stock has only 3 podcast exposure days, while TSLA leads with 1,037. Figure A.5 illustrates this concentration. This raises the concern that the baseline results may be driven by a small number of highly visible stocks.

To address this concern, Table 9 re-estimates the baseline regressions excluding the 50 most-

Table 9: Robustness to Coverage Concentration

	Full Sample		Excl. Top 50		Top 50 Only	
	Coef	t	Coef	t	Coef	t
Panel A: Market Outcomes						
Turnover	0.00336***	15.8	0.00427***	19.3	0.00202***	6.1
|Return|	0.00531***	23.1	0.00636***	24.3	0.00394***	10.4
Panel B: Retail Outcomes (Robintrack)						
Δ Log Holders	0.00519***	20.0	0.00596***	18.8	0.00329***	9.7
|Δ Holders|	45.325***	17.6	36.583***	13.3	48.868***	12.4
Controls	Included					
Fixed Effects	Stock + Calendar (DOW, WoM, Month, Year)					
Clustering	Stock					

Notes: Each cell reports the coefficient on the podcast-exposure indicator from the baseline regression (Equation 1). "Top 50" refers to the 50 stocks with the most podcast exposure stock-days, which account for 41.5% of total podcast coverage. "Excl. Top 50" removes these stocks from the sample. All specifications include the full control set, stock and calendar fixed effects, and stock-clustered standard errors. ***, **, and * denote significance at the 1%, 5%, and 10% levels.

covered stocks and, separately, using only these 50 stocks. The podcast-exposure effect remains statistically significant in both subsamples across all four outcomes. For turnover, absolute return, and Δ log holders, the coefficient is larger when the top 50 stocks are excluded. This pattern suggests that the main results are not driven by a handful of heavily covered names and, if anything, are slightly attenuated in the full sample.

4 Implications for Investor Portfolios

The preceding analyses establish that podcast coverage influences trading behavior, particularly among retail investors. Trading activity increases on podcast exposure days, and the response is stronger when coverage is more forward-looking, more actionable, and reaches larger audiences. These patterns are consistent with podcasts functioning as an information diffusion channel that shapes investor decisions. Because investors seem to react and act upon podcast signals, we next examine whether podcast-induced trading benefits investors or instead leads to greater trading activity without corresponding gains in portfolio performance.

This question connects to a large body of evidence showing that individual investors often trade to their own detriment. Barber and Odean (2000) document that the most active retail traders substantially underperform passive benchmarks, and subsequent work identifies attention-driven buying as a key mechanism: individual investors disproportionately buy stocks that catch their attention through media coverage, extreme returns, or unusual volume, and

these attention-driven purchases tend to underperform (Barber & Odean, 2008). More recently, Lalwani (2025) finds that stocks recommended by financial influencers on social media exhibit negative post-recommendation returns, suggesting that influencer-driven attention can lead investors into value-destroying trades.

To examine this possibility, we construct four portfolio strategies that reflect different ways in which investors might incorporate podcast signals into their trading decisions. All strategies impose $t+1$ execution: the investor observes the podcast signal on day t and trades at the open of day $t+1$, which is the earliest realistic implementation for most listeners (see Table OA.1 in the Online Appendix for detailed portfolio construction procedures). We benchmark each strategy against a passive buy-and-hold market portfolio.

The first strategy, the *podcast-tilted market portfolio*, captures a mild form of attention-driven overweighting. The investor holds the value-weighted market portfolio as a baseline but tilts weights toward stocks with stronger podcast signals, reflecting the well-documented tendency of retail investors to overweight stocks that are salient in the media (Barber & Odean, 2008; Focke et al., 2020). A composite signal is constructed as the mean of standardized LLM sentiment, conviction, listener score, and forward-looking score (all normalized to zero mean and unit variance on podcast exposure days, zero otherwise). Adjusted weights are $w_{i,t} = w_{i,t}^{\text{mkt}} \times (1 + \lambda \times \text{Signal}_{i,t})$ with $\lambda = 0.5$, renormalized to sum to one each day. Stocks without podcast coverage retain their exact market weight, keeping the portfolio diversified and turnover moderate.

The second strategy, the *cumulative belief portfolio*, simulates a retail investor who listens to podcasts over time and gradually builds persistent views about stocks. Rather than reacting to individual episodes, this investor forms a running belief based on an exponentially weighted moving average (EWMA) of past LLM sentiment scores with a 60-day half-life. The portfolio holds only stocks with positive cumulative belief in equal weights, rebalanced monthly. This design is motivated by models of gradual information diffusion in which investors form beliefs through repeated exposure to signals rather than through a single observation (Chen et al., 2014; Hong & Stein, 1999). It matches how an engaged podcast listener could actually invest: building up conviction gradually and acting on accumulated impressions rather than individual episodes.

The third strategy is the *long–short podcast strategy*. Stocks are sorted on the podcast signal and the investor takes long and short positions with $t+1$ execution. This strategy serves

primarily as a benchmark. It requires frequent rebalancing and short selling, which is unrealistic for most podcast listeners, but it provides a useful reference for how much, if any, of the podcast signal can be captured under a realistic implementation delay.

The fourth strategy, the *contrarian podcast portfolio*, tests whether investors can improve outcomes by doing the opposite of what podcasts recommend. The documented optimism bias of podcast hosts, which is a 3:1 buy-to-sell ratio (Section 2.2), raises the possibility that following podcast recommendations leads to systematically poor outcomes, and that a contrarian approach might fare better. Lalwani, 2025 provide evidence that financial influencer recommendations are negatively correlated with subsequent returns. Also, media-driven attention can create temporary overvaluations and can be followed by reversal effects (Hillert et al., 2014). Construction is identical to the cumulative belief portfolio but with an inverted signal: the investor holds stocks that have been consistently negatively covered.[11]

Table 10: Performance of Podcast-Based Investment Strategies

Strategy	Ann. Return (%)	Ann. Vol. (%)	Sharpe Ratio	Days
Buy & Hold (Market)	13.1	19.4	0.63	1,635
Podcast-Tilted Market $(t+1)$	14.8	19.0	0.73	1,634
Cumulative Belief $(t+1)$	10.5	19.6	0.49	1,614
Long–Short $(t+1)$	−6.3	29.5	−0.24	1,082
Contrarian Podcast $(t+1)$	4.0	21.7	0.15	1,614

Notes: Annualized performance statistics for podcast-based trading strategies over January 2016 through June 2022. All strategies execute at $t+1$: the investor observes the podcast signal on day t and trades at the open of day $t+1$. The podcast-tilted market adjusts value-weighted market portfolio weights by a composite podcast signal ($\lambda = 0.5$). The cumulative belief portfolio holds stocks with positive exponentially weighted moving average sentiment (60-day half-life), equal-weighted, monthly rebalanced. The long–short strategy sorts stocks into terciles on daily LLM sentiment and takes long and short positions. The contrarian portfolio inverts the cumulative belief signal. The Sharpe ratio uses the average daily risk-free rate annualized over the sample period. Descriptive implementation characteristics of the underlying strategies are reported in Table OA.2 and variables are defined in Table OA.1 in the Online Appendix.

Table 10 reports the results. None of the four podcast-based strategies clearly improves on passive investing. The delayed long–short strategy performs poorly, with an annualized return of −6.3% and a Sharpe ratio of −0.24, indicating that the podcast signal does not contain exploitable return-predictive information at realistic implementation horizons. This is consistent with information from publicly available media being incorporated into prices too quickly for listeners to profit from it.

Among the three more realistic strategies, the podcast-tilted market portfolio earns 14.8%

[11] Table OA.2 in the Online Appendix reports descriptive implementation characteristics of these strategies, including rebalancing frequency, the share of trading days with active positions, portfolio breadth, and, where applicable, the average size of the long and short legs. These descriptives show that the podcast-tilted market portfolio remains broadly diversified, whereas the long–short strategy is substantially more concentrated.

annualized with a Sharpe ratio of 0.73, compared with 13.1% and 0.63 for the buy-and-hold market portfolio. This modest (and statistically insignificant) difference would likely narrow further once additional turnover and monitoring costs are taken into account, which are lower in a simple buy-and-hold benchmark. The cumulative belief portfolio underperforms the market, earning 10.5% annualized with a Sharpe ratio of 0.49. This is notable because it represents the most behaviorally plausible implementation: a listener who gradually forms views based on repeated exposure to podcast content and invests accordingly. Despite this intuitive appeal, the resulting portfolio is inferior to simply holding the market. The contrarian podcast portfolio performs worst among the long-only implementations, with 4.0% annualized return and a Sharpe ratio of 0.15. Betting against the podcast consensus does not improve outcomes, suggesting that the optimism bias of podcast hosts does not create a straightforward contrarian opportunity.

Figure A.6 plots the cumulative value of $1 invested in the passive market benchmark and in each podcast-based strategy over the sample period. The figure reinforces that any outperformance of the podcast-tilted market portfolio is modest and episodic, while the cumulative belief, contrarian, and especially delayed long–short strategies fail to generate sustained value relative to passive investing. Figure A.7 places these strategies in mean-variance space relative to the passive market and the estimated efficient frontier. None of the $t+1$ podcast-based strategies lies on or close to the frontier. The delayed long–short strategy is far below the market portfolio, with negative returns and substantial volatility. The cumulative belief and contrarian strategies also lie below the market, offering lower returns for comparable or higher risk. Even the podcast-tilted market portfolio, which retains broad diversification by construction, offers at most a limited improvement over passive investing.

Taken together, these results suggest that podcast-influenced trading is unlikely to improve investor welfare. While the preceding sections document that podcasts influence how investors trade, the portfolio evidence indicates that this influence does not translate into better outcomes. This finding is consistent with the broader pattern identified by Barber and Odean (2000) and Barber and Odean (2013): media-driven attention induces trading activity, but the additional trading does not compensate investors for its costs. In the context of podcasts, the mechanism is straightforward. Podcast content is publicly available and widely consumed. By the time a listener hears a discussion, processes it, and places a trade, the information has already been incorporated into prices to the extent that it is price-relevant. The result is increased trading without corresponding gains, a pattern that is detrimental to investor portfolios particularly

when transaction costs are taken into account.

5 Conclusion

This paper shows that investing podcasts are an economically relevant channel through which public information is disseminated and incorporated into financial markets. Using a novel dataset of time-stamped podcast discussions, we document that podcast coverage is associated with immediate and persistent increases in trading activity, particularly among retail investors. We further show that investor responses to earnings announcements and M&A announcements are substantially stronger when these events are accompanied by podcast coverage, consistent with podcasts accelerating the diffusion of public information and increasing its salience.

At the same time, trading strategies that approximate how investors might act on podcast content do not generate superior returns. This suggests that podcasts do not provide an exploitable informational advantage, but rather influence how and when investors react to information that is already publicly available.

This paper has some limitations. First, although the evidence is consistent with an effect of podcast coverage on trading behavior, we cannot fully rule out that podcast discussions are timed around periods of elevated investor attention or underlying news. While we address this concern using event-time patterns, content heterogeneity, and variation in podcast release schedules, our setting does not provide a fully exogenous source of variation. The results should therefore be interpreted with appropriate caution.

Second, our measures of podcast content are based on text classification methods applied to transcribed audio, which may introduce measurement error. We mitigate this concern by employing multiple alternative measures and showing that the results are robust across specifications. Finally, the analysis focuses on a selected set of English-language investing podcasts over a specific sample period, which may limit the generalizability of the findings to other forms of financial media or different time periods.

These findings have implications for both investors and financial markets. For investors, they suggest that trading in response to podcast content may increase activity and associated costs without improving portfolio performance. For financial markets, they highlight the role of podcasts as a growing medium through which public information reaches a broader set of investors. More broadly, our results highlight the role of media-driven information diffusion in

retail trading and price formation.

References

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259–1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Barber, B. M., Huang, X., Odean, T., & Schwarz, C. (2022). Attention-induced trading and returns: Evidence from robinhood users. *The Journal of Finance*, *77*(6), 3141–3190. https://doi.org/10.1111/jofi.13183

Barber, B. M., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can investors profit from the prophets? security analyst recommendations and stock returns. *The Journal of Finance*, *56*(2), 531–563. https://doi.org/10.1111/0022-1082.00336

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, *55*(2), 773–806. https://doi.org/10.1111/0022-1082.00226

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Barber, B. M., & Odean, T. (2013). The behavior of individual investors. *Handbook of the Economics of Finance*, *2*, 1533–1570. https://doi.org/10.1016/B978-0-44-459406-8.00022-6

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *The Review of Financial Studies*, *30*(9), 3009–3047. https://doi.org/10.1093/rfs/hhx031

Blankespoor, E., Dehaan, E., Wertz, J., & Zhu, C. (2019). Why do individual investors disregard accounting information? the roles of information awareness and acquisition costs. *Journal of Accounting Research*, *57*(1), 53–84.

Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. (2021). Tracking retail investor activity. *The Journal of Finance*, *76*(5), 2249–2305. https://doi.org/10.1111/jofi.13033

Bozanic, Z., Roulstone, D. T., & Van Buskirk, A. (2018). Management earnings forecasts and other forward-looking statements. *Journal of Accounting and Economics*, *65*(1), 1–20. https://doi.org/10.1016/j.jacceco.2017.08.008

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Chae, J. (2005). Trading volume, information asymmetry, and timing information. *The Journal of Finance*, *60*(1), 413–442. https://doi.org/10.1111/j.1540-6261.2005.00734.x

Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, *27*(5), 1367–1403. https://doi.org/10.1093/rfs/hhu001

Chordia, T., Subrahmanyam, A., & Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, *59*(1), 3–32. https://doi.org/10.1016/S0304-405X(00)00080-5

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, *100*(5), 1261–1293. https://doi.org/10.1086/230638

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Eaton, G. W., Green, T. C., Roseman, B. S., & Wu, Y. (2022). Retail trader sophistication and stock market quality: Evidence from brokerage outages. *Journal of Financial Economics*, *146*(2), 502–528. https://doi.org/10.1016/j.jfineco.2022.08.002

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *the Journal of Finance*, *66*(1), 67–97.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Focke, F., Ruenzi, S., & Ungeheuer, M. (2020). Advertising, attention, and financial markets. *The Review of Financial Studies*, *33*(10), 4676–4720. https://doi.org/10.1093/rfs/hhz142

Hillert, A., Jacobs, H., & Müller, S. (2014). Media makes momentum. *The Review of Financial Studies*, *27*(12), 3467–3501. https://doi.org/10.1093/rfs/hhu061

Hirshleifer, D., Peng, L., & Wang, Q. (2025). News diffusion in social networks and stock market reactions. *The Review of Financial Studies*, *38*(3), 883–937.

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1-3), 337–386.

Hong, H., Kubik, J. D., & Stein, J. C. (2004). Social interaction and stock-market participation. *The Journal of Finance*, *59*(1), 137–163.

Hong, H., Kubik, J. D., & Stein, J. C. (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *The Journal of Finance*, *60*(6), 2801–2824.

Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance*, *55*(1), 265–295. https://doi.org/10.1111/0022-1082.00206

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, *54*(6), 2143–2184. https://doi.org/10.1111/0022-1082.00184

Huang, A. H. (2022). Classification of forward-looking statements [Available at allenhuang.org].

Huang, A. H., Wang, H., & Yang, Y. (2023). Finbert: A large language model for extracting information from financial text. *Contemporary Accounting Research*, *40*(2), 806–841. https://doi.org/10.1111/1911-3846.12832

Jiang, G. J., & Zhu, K. X. (2017). Information shocks and short-term market underreaction. *Journal of Financial Economics*, *124*(1), 43–64. https://doi.org/10.1016/j.jfineco.2016.06.006

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Lalwani, V. (2025). Finfluencer recommendations. *Economics Letters*, *255*, 112511. https://doi.org/10.1016/j.econlet.2025.112511

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, *50*(2), 495–540. https://doi.org/10.1111/j.1475-679X.2012.00450.x

Levi, S., & Zhang, X.-J. (2015). Asymmetric decrease in liquidity trading before earnings announcements and the announcement return premium. *Journal of Financial Economics*, *118*(2), 383–398. https://doi.org/10.1016/j.jfineco.2015.08.003

Li, F. (2010). The information content of forward-looking statements in corporate filings—a naive bayesian machine learning approach. *Journal of Accounting Research*, *48*(5), 1049–1102. https://doi.org/10.1111/j.1475-679X.2010.00382.x

Liu, B., Wang, H., Yu, J., & Zhao, S. (2020). Time-varying demand for lottery: Speculation ahead of earnings announcements. *Journal of Financial Economics*, *138*(3), 789–817. https://doi.org/10.1016/j.jfineco.2020.06.016

Livnat, J., & Mendenhall, R. R. (2006). Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research*, *44*(1), 177–205. https://doi.org/10.1111/j.1475-679X.2006.00196.x

Llama Team, AI @ Meta. (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*. https://doi.org/10.48550/arXiv.2407.21783

Lo, A. W., & Wang, J. (2000). Trading volume: Definitions, data analysis, and implications of portfolio theory. *The Review of Financial Studies*, *13*(2), 257–300. https://doi.org/10.1093/rfs/13.2.257

Lopez-Lira, A., & Tang, Y. (2023). Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619*. https://doi.org/10.48550/arXiv.2304.07619

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Marshall, N. T., Wegner, J., & Zechman, S. L. (2025). From mic to market: An empirical analysis of ceo podcast appearances [SSRN 4104922]. *Working Paper*. https://doi.org/10.2139/ssrn.4104922

Muslu, V., Radhakrishnan, S., Subramanyam, K. R., & Lim, D. (2015). Forward-looking md&a disclosures and the information environment. *Management Science*, *61*(5), 931–948. https://doi.org/10.1287/mnsc.2014.1921

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, *36*(4), 859–866. https://doi.org/10.1111/j.1745-9125.1998.tb01268.x

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. *Proceedings of the 40th International Conference on Machine Learning*, *202*, 28492–28518. https://doi.org/10.48550/arXiv.2212.04356

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Sinha, A., Kedas, S., Kumar, R., & Malo, P. (2019). Buy, sell or hold: Entity-aware classification of business news [W.P. No. 2019-04-02]. *IIMA Working Paper*.

So, E. C., & Wang, S. (2014). News-driven return reversals: Liquidity provision ahead of earnings announcements. *Journal of Financial Economics*, *114*(1), 20–35. https://doi.org/10.1016/j.jfineco.2014.06.009

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*(3), 1437–1467. https://doi.org/10.1111/j.1540-6261.2008.01362.x

Welch, I. (2022). The wisdom of the robinhood crowd. *The Journal of Finance*, *77*(3), 1489–1527. https://doi.org/10.1111/jofi.13128

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *The Journal of Finance*, *51*(1), 137–167. https://doi.org/10.1111/j.1540-6261.1996.tb05205.x

Appendix

Table A.1: Podcast Sample

1	7:31 AM	48	Pounding The Table
2	AJ Bell Money & Markets	49	Quoth the Raven
3	Ask The Compound	50	Rangeley Capital Podcast
4	Barron's Streetwise	51	Rebel Traders Podcast
5	Behind the Balance Sheet	52	RiskReversal Pod
6	Ledger Cast	53	Rule Breaker Investing
7	Bloomberg Businessweek	54	Squawk on the Street
8	Bloomberg Intelligence	55	Stansberry Investor Hour
9	Bloomberg Surveillance	56	Stock Club
10	BloombergQuint	57	Stock Market Options Trading
11	Business Breakdowns	58	Stock Moe Money Show
12	Capital Ideas Podcast	59	Stock Trading University
13	Cash Daddies	60	Switched On
14	Chat With Traders	61	Talking Trading
15	Chit Chat Stocks	62	Tesla Daily
16	CNBC's "Fast Money"	63	The What is Money? Show
17	CNBC's "Options Action"	64	The Acquirers Podcast
18	Desire To Trade Podcast	65	The Business Brew
19	Devils and Details	66	The Canadian Investor
20	Dumb Money Live	67	The Compound and Friends
21	Exchanges	68	The Dividend Cafe
22	Eye On The Market	69	The Exchange
23	Focused Compounding	70	The HC Commodities Podcast
24	Forward Guidance	71	The Intelligent Investing Podcast
25	Futures Radio Show	72	The Investing City Podcast
26	FYI - For Your Innovation	73	The Jay Martin Show
27	Halftime Report	74	The Joseph Carlson Show
28	Hedgeye Podcasts	75	The MoneyWeek Podcast
29	In Penny Stock	76	The Scoop
30	Investing With IBD	77	The SharePickers Podcast
31	Investments Unplugged	78	The Stock Trading Reality Podcast
32	Investor Cheat Code Podcast	79	The Wall Street Lab
33	INVESTOR IN THE FAMILY Radio	80	Thoughts on the Market
34	Investors' Chronicle	81	Today's Market Explained
35	InvestTalk	82	Trading Camp
36	Mad Money w/ Jim Cramer	83	Trading Nut
37	Market Maker	84	Unconfirmed
38	Market Mondays	85	Value Investing with Legends
39	Mining Stock Daily	86	Wake Up to Money
40	Money Metals Market Wrap	87	Wall St For Main St
41	Motley Fool Money	88	Wall Street Unplugged
42	MRKT Matrix	89	What Goes Up
43	Not Investment Advice	90	Worldwide Exchange
44	Notes on the Week Ahead	91	WorldWide Markets
45	Odd Lots	92	Yet Another Value Podcast
46	Patrick Boyle On Finance	93	Zacks Market Edge
47	Penny Lane Podcast		

Notes: This table lists all 93 podcasts in our sample, sorted alphabetically. Podcasts are selected from the Listen Notes database by filtering for English-language shows tagged "Investing" and manually screening the 1,000 most popular podcasts for equity-focused content that regularly discusses individual stocks. One podcast (*Motley Fool Money*) appears under two separate feeds, yielding 94 unique podcast IDs.

Table A.2: Specification Curve Design Dimensions

Dimension	Categories	n
Exposure measure	Podcast exposure, log LS-wtd mentions, log snippets, log episodes, log shows, abs. sentiment, conviction, log words	8
Outcome variable	Turnover, |return|, Δ log holders, $\|\Delta$ holders$\|$	4
Fixed effects	Stock + Calendar, Stock only	2
Control set	Baseline, Expanded	2
Sample	Full, Large-cap ($>$ median), Excl. microcaps ($>$ P20)	3
Winsorization	1%, 0.5%	2
Total specifications: $8 \times 4 \times 2 \times 2 \times 3 \times 2$		1,152

Notes: This table summarizes the six design dimensions varied in the specification curve analysis and the number of categories within each dimension. The full factorial combination of these choices yields 1,152 distinct specifications underlying Figure A.4.

Table A.3: Podcast Effects by Episode Release Cadence

	(1) Turnover	(2) |Return|	(3) Δ Log Hold.	(4) $\|\Delta$ Hold.$\|$
On-Schedule Mention	0.0030^{***}	0.0046^{***}	0.0049^{***}	46.88^{***}
	(0.0002)	(0.0003)	(0.0003)	(3.13)
	[13.1]	[17.5]	[15.5]	[15.0]
Off-Schedule Mention	0.0036^{***}	0.0057^{***}	0.0054^{***}	44.82^{***}
	(0.0002)	(0.0003)	(0.0003)	(2.65)
	[16.0]	[21.4]	[17.2]	[16.9]
Difference (On $-$ Off)	-0.0006^{***}	-0.0011^{***}	-0.0005	2.07
	[-3.5]	[-3.8]	[-1.4]	[0.8]
Controls	Included			
Fixed Effects	Stock + Calendar (DOW, WoM, Month, Year)			
Clustering	Stock			
N	3,524,669	3,524,669	1,151,131	1,151,131
On-Schedule stock-days	19,022 (42% of podcast exposure days)			
Off-Schedule stock-days	25,883 (58% of podcast exposure days)			

Notes: This table decomposes the podcast-exposure effect by whether the episode was released on its show's regular publication day. For each podcast show, we compute the empirical distribution of release days across the week and identify the show's peak release day. A mention is classified as *on-schedule* if the episode aired on the show's peak weekday and as *off-schedule* otherwise. Both indicators enter the regression simultaneously; the reference category is non-exposure days. On-schedule mentions proxy for supply-driven podcast exposure, where the episode airs because it was scheduled to air, not because of breaking news about a particular stock. The difference row reports the test of equality of the two coefficients based on the joint covariance matrix. Controls include lagged return, lagged turnover, lagged absolute return, and log market capitalization. Standard errors clustered at the stock level are in parentheses; t-statistics in brackets. ***, **, and * denote significance at the 1%, 5%, and 10% levels.

Table A.4: Incremental Podcast Effects Around M&A Announcements

	Turnover (1)	|Return| (2)	Δ Log Hold. (3)	|Δ Hold.| (4)
M&A Window (β_1)	0.00058**	0.00322***	0.00196***	−1.685
	(0.00024)	(0.00042)	(0.00075)	(2.808)
	[2.4]	[7.6]	[2.6]	[-0.6]
Podcast Exposure (β_2)	0.00320***	0.00503***	0.00503***	45.062***
	(0.00021)	(0.00022)	(0.00025)	(2.590)
	[15.5]	[23.0]	[19.9]	[17.4]
M&A × Podcast (β_3)	0.00994***	0.01218***	0.01456***	52.935***
	(0.00188)	(0.00239)	(0.00364)	(17.645)
	[5.3]	[5.1]	[4.0]	[3.0]
Implied effects on M&A days:				
M&A without podcast (β_1)	0.00058	0.00322	0.00196	−1.685
M&A with podcast ($\beta_1 + \beta_2 + \beta_3$)	0.01372	0.02043	0.02155	96.312
Difference ($\beta_2 + \beta_3$)	+0.01314	+0.01720	+0.01959	+97.997
Wald F-statistic	47.6***	51.1***	28.7***	31.4***
Controls	Baseline set + Log Deal Value + |Announcement Return|			
Fixed Effects	Stock + Calendar (DOW, WoM, Month, Year)			
Clustering	Stock			

Notes: Each column reports a single regression analogous to Equation (4), with M&A announcement window (±1 trading day, from SDC Platinum) replacing the earnings indicator. Deal-level controls include log deal value and the absolute 1-day announcement return. Of 44,905 podcast exposure stock-days, 478 (1.1%) fall within an M&A window. The Wald F-test tests $H_0: \beta_2 + \beta_3 = 0$. Columns (3)–(4) use the Robintrack subsample. t-statistics in brackets. ***, **, and * denote significance at the 1%, 5%, and 10% levels. See Table OA.1 in the Online Appendix for detailed variable definitions.

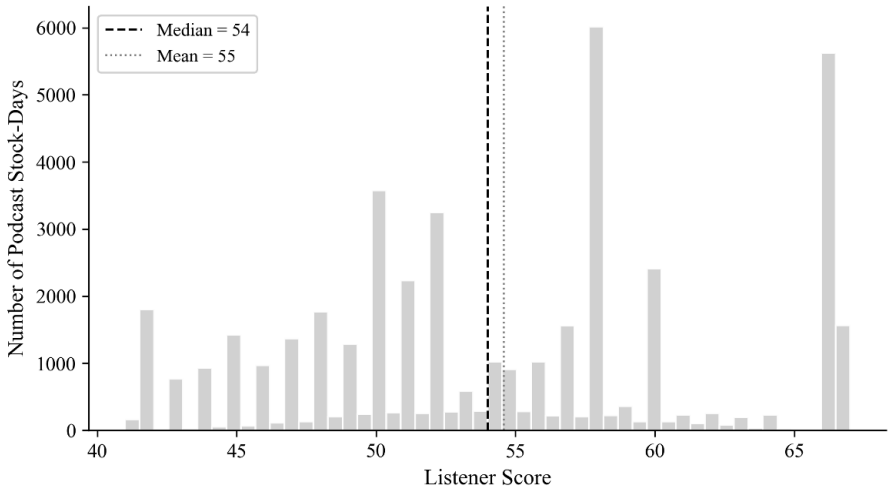


Figure A.1: Distribution of Listener Scores Across Podcast Stock-Days

Notes: Histogram of mean listener scores across 44,905 podcast exposure stock-days. The dashed line marks the median (54), the dotted line the mean (55). Listener scores range from 41 to 67 and capture podcast audience reach on a proprietary index.

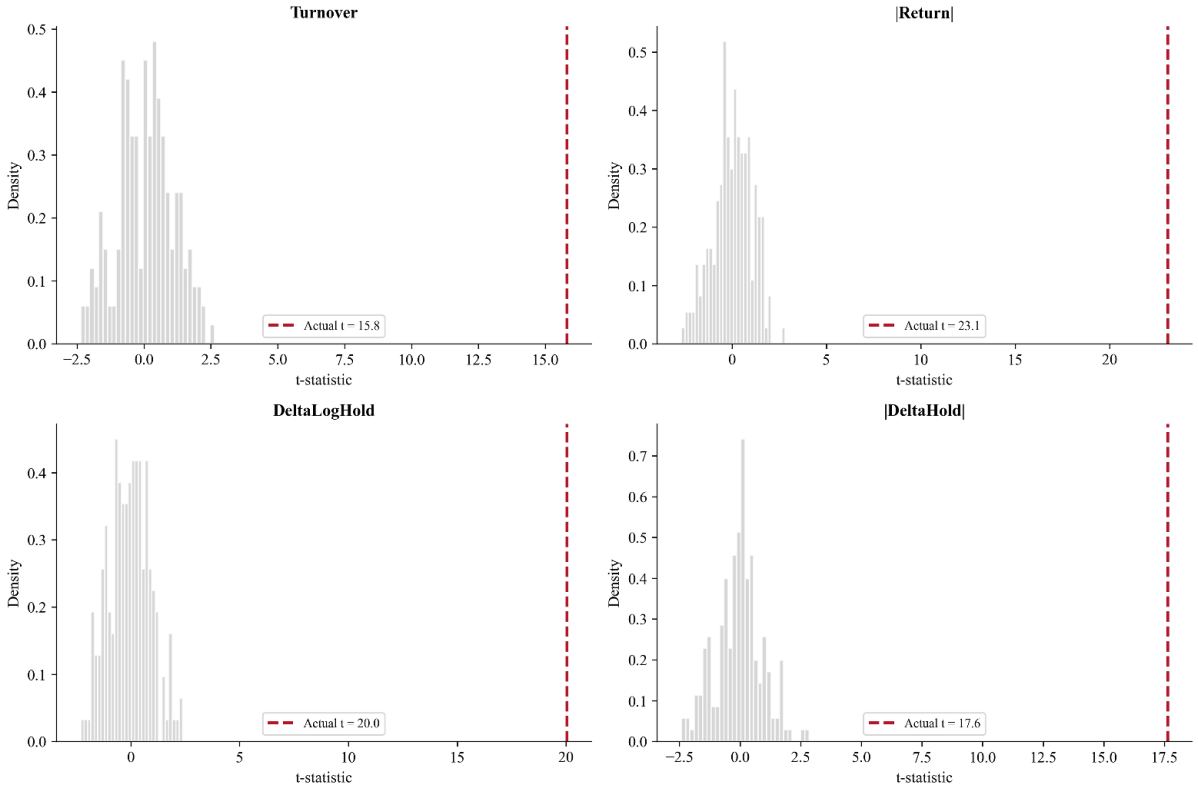


Figure A.2: Permutation Test: Null Distribution of t-Statistics (All Outcomes)

Notes: Histograms of t-statistics from 200 within-stock random reassignments of podcast exposure days. Vertical dashed lines indicate the actual t-statistics. Permutation $p < 0.001$ for all four outcomes.

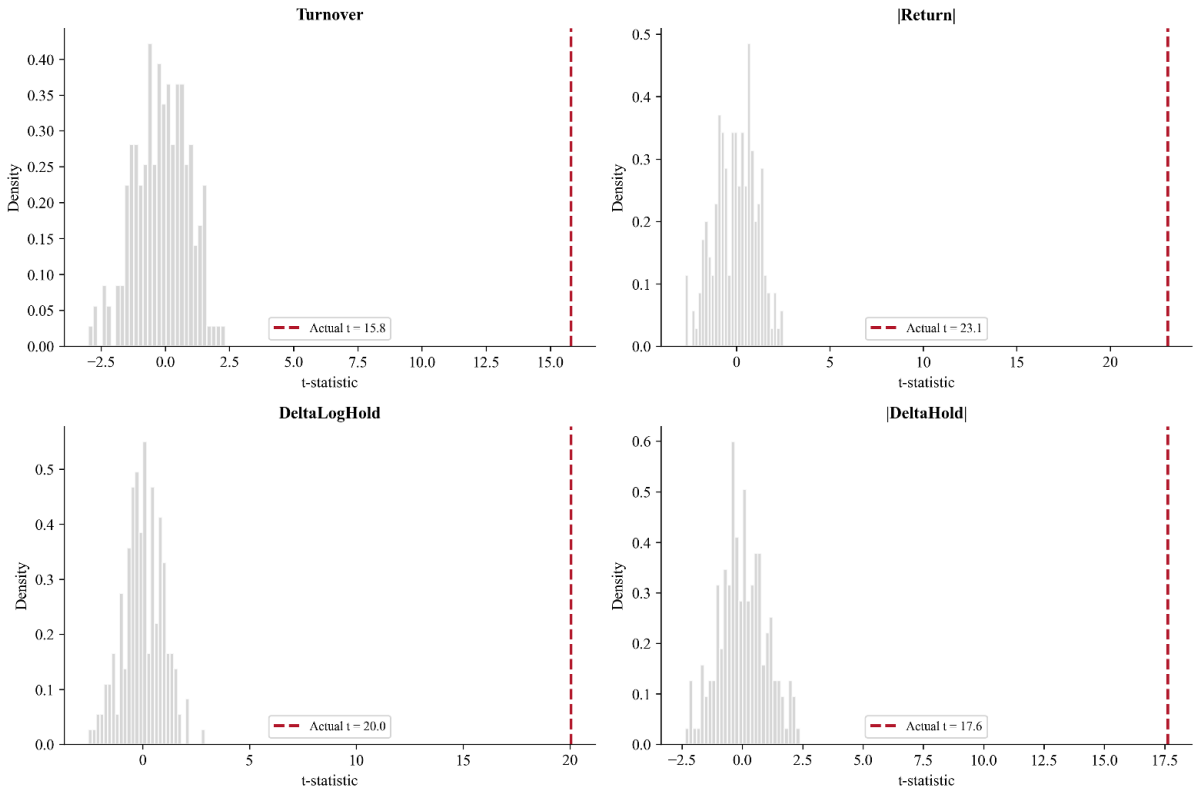


Figure A.3: Random Date Placebo: Null Distribution (All Outcomes)

Notes: Histograms of t-statistics from 200 random calendar-date assignments, preserving stock-level podcast frequency. Vertical dashed lines indicate the actual t-statistics. Placebo $p < 0.001$ for all four outcomes.

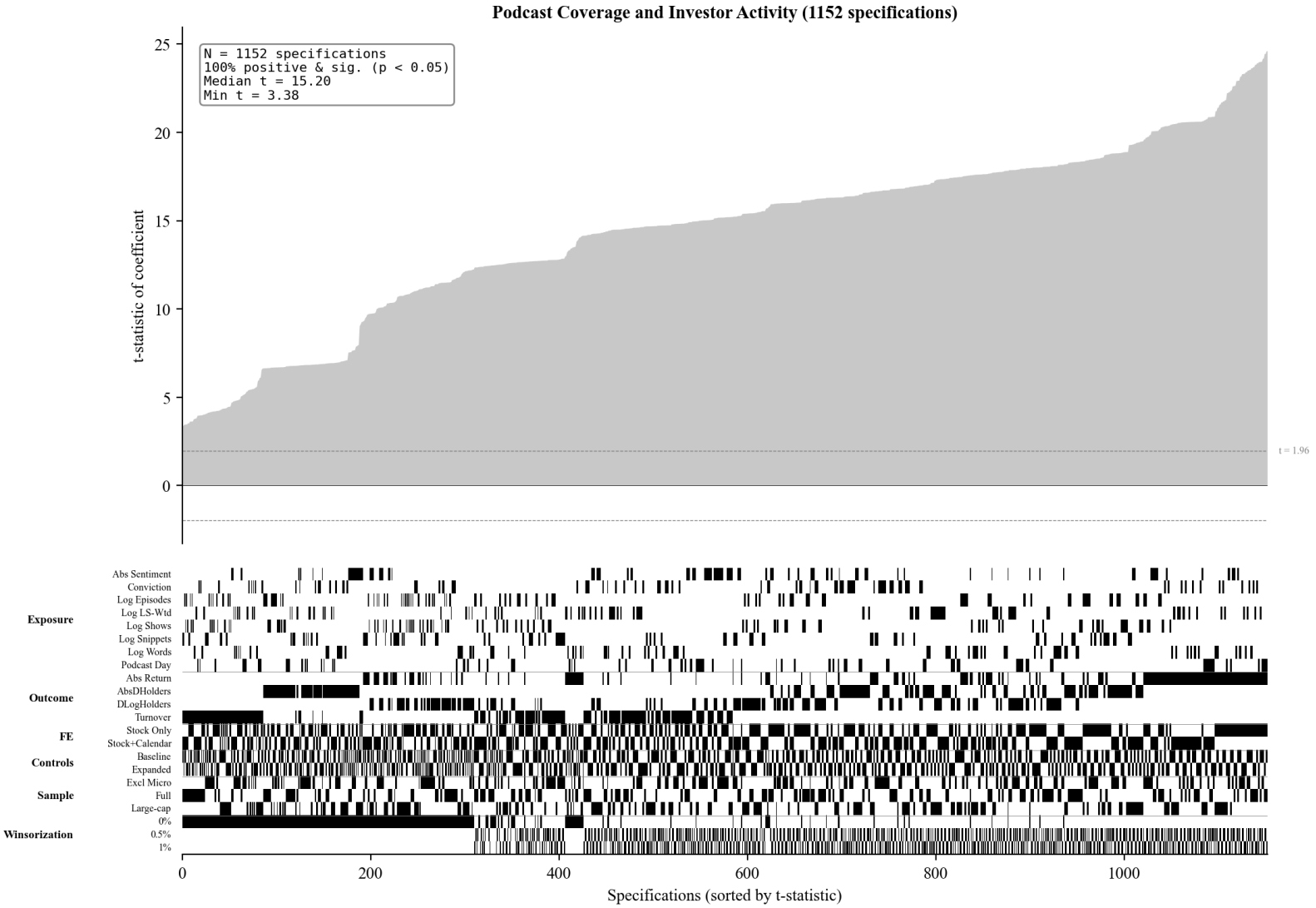


Figure A.4: Specification Curve: Podcast Coverage and Investor Response

Notes: Each point represents one of 1,152 specifications varying across six design dimensions (Table A.2). The upper panel plots t-statistics sorted by magnitude; the lower panel indicates which design choices are active for each specification. All specifications use stock-clustered standard errors.

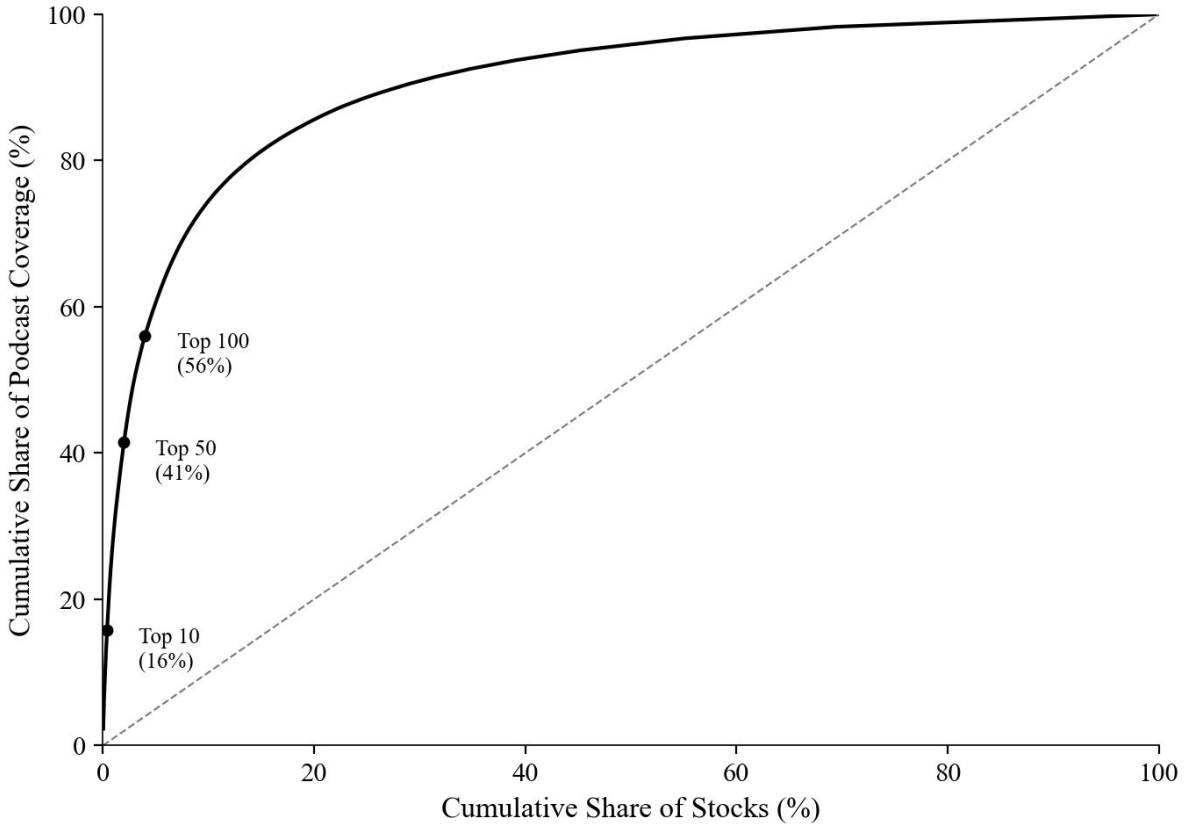


Figure A.5: Podcast Coverage Concentration Across Stocks

Notes: Cumulative share of podcast exposure stock-days as a function of cumulative share of stocks (ranked by coverage frequency). The dashed line represents perfect equality. The top 10 stocks account for 16% of all podcast exposure stock-days, the top 50 for 41%, and the top 100 for 56%. 2,500 unique stocks receive at least one podcast mention.

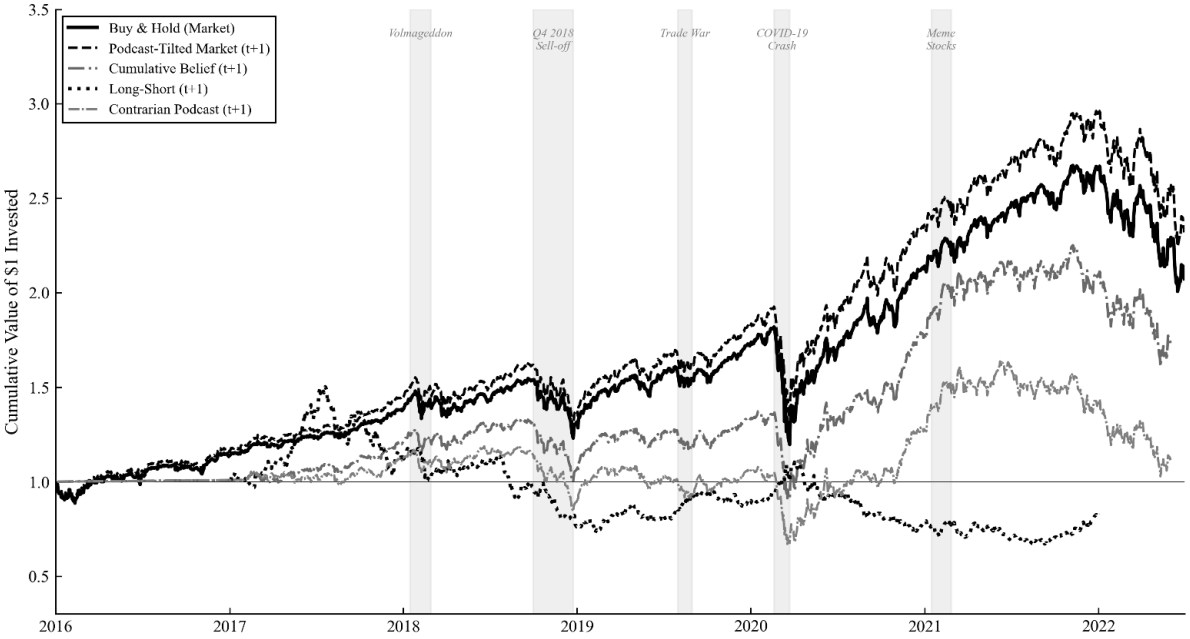


Figure A.6: Cumulative Performance of Podcast-Based Investment Strategies

Notes: This figure plots the cumulative value of $1 invested in the passive market benchmark and in each podcast-based strategy from January 2016 through June 2022. All podcast-based strategies use $t+1$ execution, so that the investor observes the podcast signal on day t and trades at the open of day $t+1$. The buy-and-hold benchmark is the value-weighted CRSP market portfolio. The long–short series represents the cumulative return on the self-financing long–short strategy, normalized to one at the beginning of the sample for comparability. Gray bands mark five salient market events: Volmageddon (Jan–Feb 2018), the Q4 2018 sell-off (Oct–Dec 2018), the U.S.–China trade-war escalation (Aug 2019), the COVID-19 crash (Feb–Mar 2020), and the meme-stock episode (Jan 2021).

Table A.5: Alternative Sentiment Measures

	Turnover		Δ Log Holders	
Measure	Coef	t	Coef	t
LLM Sentiment	-0.00086^{***}	-5.0	-0.00210^{***}	-6.1
FinBERT	-0.00131^{***}	-5.7	-0.00355^{***}	-6.2
LM Dictionary	-0.02495^{***}	-6.0	-0.03628^{***}	-3.8
TF-IDF	-0.01504^{***}	-11.1	-0.02009^{***}	-8.0

Notes: This table tests whether the sentiment-based trading results are robust to alternative sentiment measures. Each row replaces the primary stock-day LLM sentiment score with a different proxy derived from the same podcast snippets: FinBERT, the Loughran–McDonald dictionary, and a TF-IDF prototype-based sentiment measure. Reported are the coefficient on the sentiment variable and its t-statistic from the baseline regression. Columns (1)–(2) use turnover as the dependent variable; columns (3)–(4) use Δ log holders in the Robintrack subsample. All regressions include the standard controls, stock and calendar fixed effects, and stock-clustered standard errors. The coefficient remains negative and statistically significant across all alternative sentiment measures.

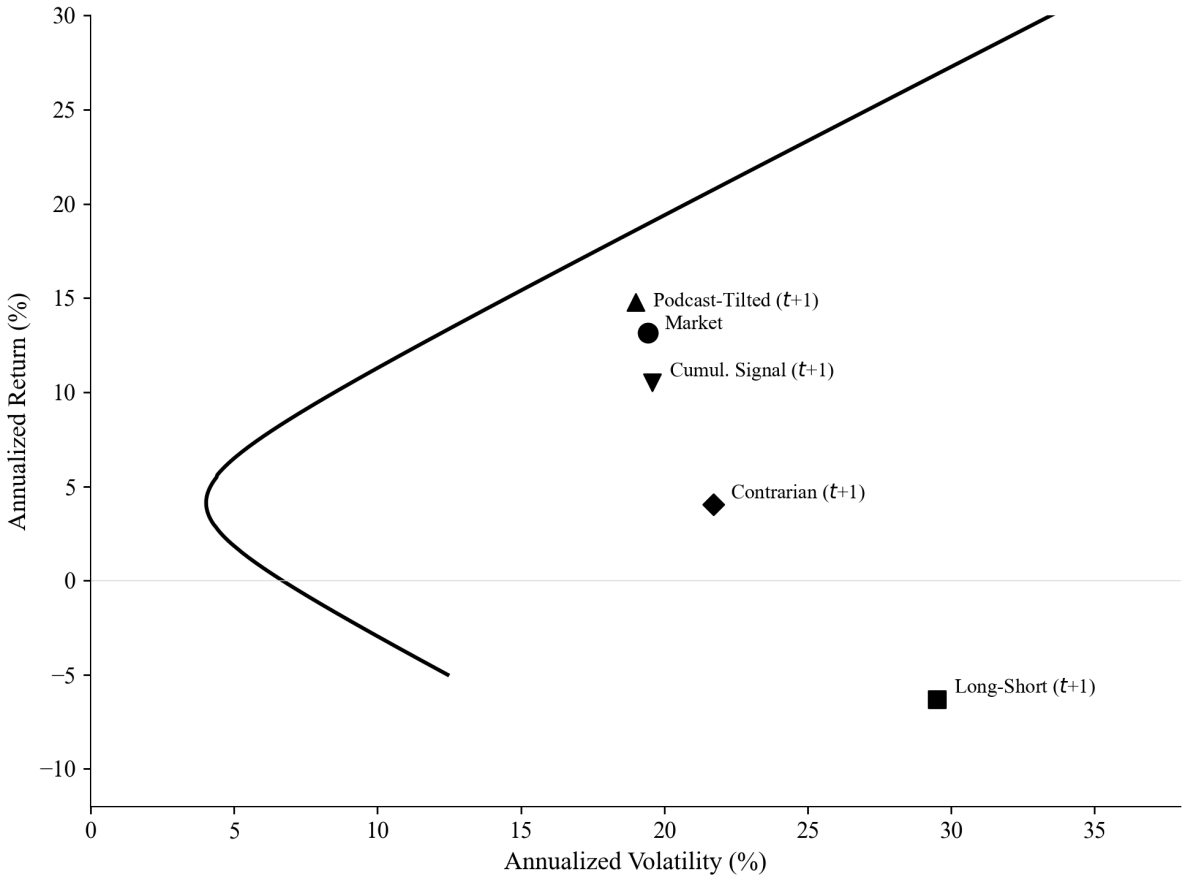


Figure A.7: Mean-Variance Positioning of Podcast Investment Strategies

Notes: Each point represents the annualized mean return and volatility of a portfolio strategy over January 2016 through June 2022. The solid curve is the efficient frontier estimated from Fama–French six factors. All podcast strategies shown in the figure use $t + 1$ execution. The market portfolio (circle) is the value-weighted CRSP market.

Online Appendix

Table OA.1: Variable Definitions and Construction Details

Variable	Description	Construction Details	References
Panel A: Podcast Data Construction			
Podcast Sample	Set of investing podcasts from which stock-specific content is extracted.	From Listen Notes, all English-language podcasts tagged "Investing" are retrieved. The 1,000 most popular shows are manually screened; a show is retained if it (i) focuses on equity investing and (ii) regularly discusses individual stocks. Yields 94 podcast IDs (93 unique titles). See Table A.1 for the full list.	—
Listener Score	Podcast audience-reach proxy; composite popularity index on a 0–100 scale from Listen Notes.	Aggregates platform-internal engagement signals (e.g., plays, estimated downloads) with external indicators (e.g., online mentions, reviews). Observed range in sample: 41–67 (median 54). Used to construct LS-weighted aggregates.	—

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
Snippet	Unit of observation: one self-contained text passage about a specific stock on a specific day.	Episode transcripts (OpenAI Whisper) are processed by Llama-3.1-8B-Instruct (temperature = 0) in a multi-pass pipeline. **Pass 1** (stock identification): a sliding window (200 words, step 150) is scored for stock references using a few-shot prompt with five labeled examples. **Pass 2** (validation): top-N candidate chunks (N=10) are re-evaluated by the LLM. **Extraction**: validated passages are extracted as verbatim text and linked to a ticker and calendar date. Yields 102,623 snippets across 4,570 tickers and 1,650 days.	Radford et al. (2023); Llama Team, AI @ Meta (2024)
Panel B: Snippet-Level Content Measures			
LLM Sentiment	Primary sentiment measure; continuous score on $[-1, +1]$.	A standardized zero-shot prompt instructs the LLM to assess the directional tone of the stock discussion. Prompt (excerpt): "Score this financial podcast snippet about {ticker} on four dimensions. 1. SENTIMENT: How positive/negative about this stock? -1.0 = very negative/bearish, 0.0 = neutral, $+1.0$ = very positive/bullish." The sentiment dimension is extracted from the combined four-dimension scoring call (see below). Used as the primary trading signal in portfolio construction.	Lopez-Lira and Tang (2023)
FinBERT Sentiment	Robustness sentiment measure; continuous score.	ProsusAI/finbert, a finance-adapted BERT model. Returns class probabilities for positive, negative, and neutral. The continuous score is $P(\text{positive}) - P(\text{negative})$. Correlation with LLM Sentiment: $\rho = 0.42$.	Huang et al. (2023)

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
LM Dictionary Sentiment	Dictionary-based robustness sentiment measure.	Net sentiment tone from the Loughran and McDonald (2011) financial dictionary. Computed as (positive word count − negative word count) / total words. Correlation with FinBERT: $\rho = 0.40$.	Loughran and McDonald (2011)
TF-IDF Sentiment	Prototype-based robustness sentiment measure.	Each snippet is converted to a TF-IDF vector. Sentiment is measured as cosine similarity to positive minus negative prototype vectors. Prototypes are constructed from 10 seed texts (5 positive, 5 negative) defined in a configuration file. The vector-space approach follows Salton and Buckley (1988).	Salton and Buckley (1988)
Forward-Looking (LLM)	Primary forward-looking measure; continuous score on $[0, 1]$.	Extracted from the combined LLM scoring call. Prompt dimension: "2. FORWARD_LOOKING: Does snippet contain forward-looking statements? 0.0 = purely backward-looking, 1.0 = strongly forward-looking."	Bozanic et al. (2018); Li (2010)
FinBERT-FLS	Robustness forward-looking measure.	yiyanghkust/finbert-fls, trained on 3,600 labeled MD&A sentences. Returns probabilities for Specific FLS, Non-Specific FLS, and Not-FLS. Continuous score: $P(\text{Specific FLS}) + P(\text{Non-Specific FLS})$.	Muslu et al. (2015); Bozanic et al. (2018); yya518.github.io/finbert
Li FLS Dictionary	Dictionary-based robustness forward-looking measure.	Counts forward-looking keywords from 16 narrow category groups plus a broad list, following Li (2010). Expressed as share of forward-looking words in total words.	Li (2010)

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
Actionability	Primary actionability measure; TF-IDF prototype similarity score.	Measures cosine similarity of the snippet's TF-IDF vector to four reference profiles (buy, sell, hold, none). Actionability is defined as the maximum similarity to any of the three recommendation prototypes minus the similarity to the no-recommendation prototype. Largely orthogonal to sentiment ($\rho < 0.12$).	Salton and Buckley (1988)
Recommendation Score	Directional recommendation strength; continuous on $[-1, +1]$.	A dedicated LLM prompt instructs the model to determine whether the snippet contains a concrete investment recommendation and to classify it as BUY, SELL, HOLD, or NONE. The model returns a categorical label and a continuous directional score.	Womack (1996); Barber et al. (2001); Sinha et al. (2019)
Buy/Sell/Hold Signal	Categorical recommendation indicator (BUY, SELL, HOLD, NONE).	Extracted from the same dedicated recommendation LLM prompt. At the stock-day level, indicator variables equal one if at least one snippet contains the respective signal. Among podcast exposure day observations: 46.7% buy, 16.0% sell, 13.2% hold, 34.0% none.	Womack (1996); Barber et al. (2001)
Conviction (LLM)	Primary conviction measure; continuous on $[0, 1]$.	Extracted from the combined LLM scoring call. Prompt dimension: "3. CONVICTION: How strongly/confidently expressed? 0.0 = uncertain/hedging, 1.0 = confident/definitive."	Larcker and Zakolyukina (2012)
TF-IDF Conviction	Robustness conviction measure (prototype-based).	Cosine similarity of the snippet TF-IDF vector to confident vs. hedged prototype language, constructed analogously to TF-IDF Sentiment.	Salton and Buckley (1988)
LM Strong/Weak Modal	Dictionary-based conviction proxies.	Counts of strong modal words (e.g., "will," "always") and uncertainty/weak modal words (e.g., "may," "possibly") from Loughran and McDonald (2011).	Loughran and McDonald (2011)

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
Hedge-to-Boost Ratio	Custom dictionary conviction proxy.	Ratio of hedge words (31 terms, e.g., "maybe," "perhaps") to boost/intensifier words (41 terms, e.g., "definitely," "strongly"). Custom word lists; see replication code.	Larcker and Zakolyukina (2012)
Panel C: Stock-Day Podcast Aggregates			
Podcast Exposure	Binary indicator; equals one if stock i receives ≥ 1 snippet mention on day t.	Constructed during stock-day panel creation. 44,905 podcast exposure stock-days out of 3,527,519 total (1.27%).	—
LS-Weighted Mentions	Continuous podcast intensity measure.	Sum of listener scores across all snippets mentioning stock i on day t. Set to zero on non-exposure days. Log-transformed when used as exposure: $\log(1 + \text{LS-wtd mentions})$.	—
Stock-Day Sentiment	Stock-day aggregated sentiment.	Unweighted mean and listener-score-weighted mean of snippet-level LLM Sentiment for stock i on day t. Set to zero on non-exposure days. LS-weighted mean is the preferred aggregation.	Tetlock (2007)
Stock-Day Forward-Looking	Stock-day aggregated forward-looking score.	Constructed identically to stock-day sentiment, using the LLM forward-looking score.	—
Stock-Day Conviction	Stock-day aggregated conviction score.	Constructed identically to stock-day sentiment, using the LLM conviction score.	—
Recommendation Counts	Number of buy, sell, and hold snippets per stock-day.	Sum of categorical recommendation labels across all snippets for stock i on day t.	—

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
Panel D: Outcome Variables			
Turnover	Daily share turnover; primary market-activity outcome.	Daily trading volume divided by shares outstanding, from CRSP. Winsorized at 1st and 99th percentiles.	Lo and Wang (2000); Chordia et al. (2001)
|Return|	Absolute daily return; captures magnitude of price adjustment.	Absolute value of the CRSP daily holding-period return. Winsorized at 1/99.	Da et al. (2011); Ben-Rephael et al. (2017)
Daily Return	Signed daily stock return.	CRSP daily holding-period return (`ret`). Used in portfolio evaluation. Winsorized at 1/99.	—
Δ Log Holders	Proportional net buying by Robinhood retail investors.	$\Delta \log(\text{holders}_{i,t}) = \log(\text{holders}_{i,t}) - \log(\text{holders}_{i,t-1})$, from Robintrack daily ticker-level user counts (May 2018–August 2020). Winsorized at 1/99.	Barber et al. (2022); Welch (2022); Eaton et al. (2022)
$\lvert\Delta$ Holders$\rvert$	Absolute change in Robinhood holder counts; captures retail trading intensity regardless of direction.	$\lvert\text{holders}_{i,t} - \text{holders}_{i,t-1}\rvert$, from Robintrack. Winsorized at 1/99.	Barber et al. (2022); Welch (2022); Eaton et al. (2022)
Panel E: Control Variables			
Log Market Cap	Firm size control.	$\log(\texttt{prc} \times \texttt{shrout})$ from CRSP daily. Entered as one-day lag $\log(\text{mktcap}_{i,t-1})$.	—

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
Book-to-Market (BTM)	Value/growth proxy.	`ceq`/`mkvalt` from Compustat Annual; fallback: (`seq` − `pstk`)/`mkvalt`. Matched to CRSP via the CRSP–Compustat linking table with a 6-month publication lag. Winsorized at 1/99.	Fama and French (1993)
Leverage	Financial leverage control.	(`dltt` + `dlc`)/`at` from Compustat Annual; `dlc` filled with zero if missing. Winsorized at 1/99.	—
Return on Assets (ROA)	Profitability control.	`ib`/`at` from Compustat Annual. Winsorized at 1/99.	Fama and French (2015)
Log Analyst Following	Analyst coverage control.	log(1 + `NUMEST`), where `NUMEST` is the number of analysts issuing EPS forecasts from the most recent I/B/E/S Summary Statistics record (measure = EPS, FPI = '1'). Matched to the panel via `merge_asof` on the most recent `STATPERS` date.	Hong et al. (2000)
Lagged Return	Prior-day return control.	$\texttt{ret}_{i,t-1}$ from CRSP.	—
Lagged Turnover	Prior-day trading activity control.	$\text{Turnover}_{i,t-1}$ as defined above.	—
Lagged |Return|	Prior-day absolute return control.	$\lvert\text{Return}\rvert_{i,t-1}$ as defined above.	—
Panel F: Fixed Effects			
Stock FE (α_i)	Absorbs time-invariant stock characteristics.	Implemented via within-group demeaning in OLS.	—

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
Calendar FE ($\boldsymbol{\delta}_{\text{cal}(t)}$)	Absorbs systematic calendar-time trading patterns.	Four sets of indicators: day-of-week (DOW, 5 levels), week-of-month (WoM, $\lceil \text{day}/7 \rceil$, 5 levels), month-of-year (12 levels), and year. Implemented jointly via within-group demeaning.	—
Panel G: Corporate Events and Interaction Variables			
EA Window	Earnings announcement indicator.	Equals one if stock i has an I/B/E/S-reported quarterly earnings announcement (USFIRM = 1, MEASURE = 'EPS', PDICITY = 'QTR') within ± 1 calendar day of `ANNDATS`. Constructed by expanding each announcement date by $\{-1, 0, +1\}$ days.	—
Earnings Surprise	Actual minus expected EPS.	Actual EPS minus the consensus mean analyst forecast from I/B/E/S. Winsorized at 1/99 and set to zero on non-EA days. Used as a control in the EA interaction specification (Table 5).	Livnat and Mendenhall (2006)
|Earnings Surprise|	Absolute earnings surprise.	Absolute value of Earnings Surprise. Winsorized at 1/99 and set to zero on non-EA days.	Livnat and Mendenhall (2006)
EA × Podcast	Interaction of EA window and podcast exposure.	EA Window$_{i,t}$ × Podcast Exposure$_{i,t}$. Captures the incremental podcast effect conditional on a public earnings event. Coefficient β_3 in Equation (4).	—
M&A Window	M&A announcement indicator.	Equals one if stock i has a deal announcement from SDC Platinum within ± 1 trading day of `DATEANN`. Sample restricted to SDC target tickers matched to the panel.	—
Log Deal Value	M&A deal size control.	log(deal value in $ millions) from SDC Platinum. Set to zero on non-M&A days.	—

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
|Announcement Return|	M&A price impact control.	Absolute 1-day CRSP return on the SDC announcement date. Set to zero on non-M&A days.	—
M&A × Podcast	Interaction of M&A window and podcast exposure.	M&A Window$_{i,t}$ × Podcast Exposure$_{i,t}$. Captures incremental podcast diffusion effect around M&A events.	—
Panel H: Retail vs. Market Classification			
Retail-Dominant	Quadrant indicator for retail-skewed trading.	Equals one if |Δholders| is in the top decile of the cross-sectional distribution on day t and turnover is below the top decile. Robintrack subsample only.	Boehmer et al. (2021)
Market-Dominant	Quadrant indicator for market-skewed trading.	Equals one if turnover is in the top decile on day t and |Δholders| is below the top decile.	—
Both Active	Quadrant indicator for jointly elevated trading.	Equals one if both turnover and |Δholders| are in the top decile on day t.	—
Both Quiet	Quadrant indicator for jointly low trading.	Equals one if neither turnover nor |Δholders| is in the top decile on day t.	—
Positive/Negative Podcast Exposure	Sentiment decomposition indicators.	A podcast exposure day is classified as *positive* if the stock-day mean LLM sentiment > 0.1 and *negative* if < -0.1. Both indicators enter the same regression; the threshold ensures a buffer around zero. Results robust to FinBERT and to a ± 0 cutoff.	Tetlock (2007); Kahneman and Tversky (1979)
Panel I: Alternative Podcast Exposure Measures			

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
Log Snippet Count	Volume of stock-specific coverage.	$\log(1 + \text{number of snippets}_{i,t})$. Set to zero on non-exposure days.	—
Log Episode Count	Breadth across episodes.	$\log(1 + \text{number of distinct episodes}_{i,t})$. Set to zero on non-exposure days.	—
Log Show Count	Breadth across distinct podcast shows.	$\log(1 + \text{number of distinct shows}_{i,t})$. Set to zero on non-exposure days.	—
Absolute LLM Sentiment	Extremity of podcast signal regardless of direction.	$\lvert\text{LLM Sentiment}_{i,t}\rvert$. Set to zero on non-exposure days.	—
Conviction-Weighted Signal	Conviction-adjusted sentiment.	$\text{LLM Sentiment}_{i,t} \times \text{Conviction}_{i,t}$. Set to zero on non-exposure days.	—
Log Total Words	Depth of discussion.	$\log(1 + \text{total words across all snippets}_{i,t})$. Set to zero on non-exposure days.	—
Panel J: Portfolio Construction and Evaluation			
Podcast-Tilted Market	Conservative institutional strategy (t+1).	Holds the value-weighted CRSP market portfolio with tilted weights: $w_{i,t} = w_{i,t}^{\text{mkt}} \times (1 + \lambda \times \text{Signal}_{i,t})$, $\lambda = 0.5$. Signal is the mean of standardized LLM sentiment, conviction, listener score, and forward-looking score (normalized to zero mean, unit variance on podcast exposure days; zero otherwise). Weights renormalized to sum to one daily. Non-covered stocks retain exact market weights.	—
Cumulative Belief Portfolio	Retail-oriented long-only strategy (t+1).	EWMA of past LLM sentiment with a 60-day half-life. Holds only stocks with positive cumulative belief in equal weights, monthly rebalanced.	—

Continued on next page

Table OA.1 continued

Variable	Description	Construction Details	References
Delayed Long-Short	Next-day implementation of the tercile strategy.	Identical to the long-short tercile portfolio but with t+1 execution.	—
Contrarian Podcast Portfolio	Inverted cumulative belief (t+1).	Identical to the cumulative belief portfolio but with inverted signal: holds stocks with consistently negative podcast coverage.	—
Fama-French Factors	Risk factors for portfolio evaluation.	Daily returns on MKT-RF, SMB, HML, RMW, CMA (Fama-French five factors) plus UMD (momentum), from Ken French's data library. Used for alpha estimation and efficient-frontier construction.	Fama and French (2015); Carhart (1997)
Parkinson Volatility	Intraday volatility measure (robustness).	$\sigma^2_{\text{Park},i,t} = [\log(\texttt{askhi}_{i,t}/\texttt{bidlo}_{i,t})]^2/(4\log 2)$ from CRSP daily high/low prices. Used in expanded control sets and robustness checks.	—

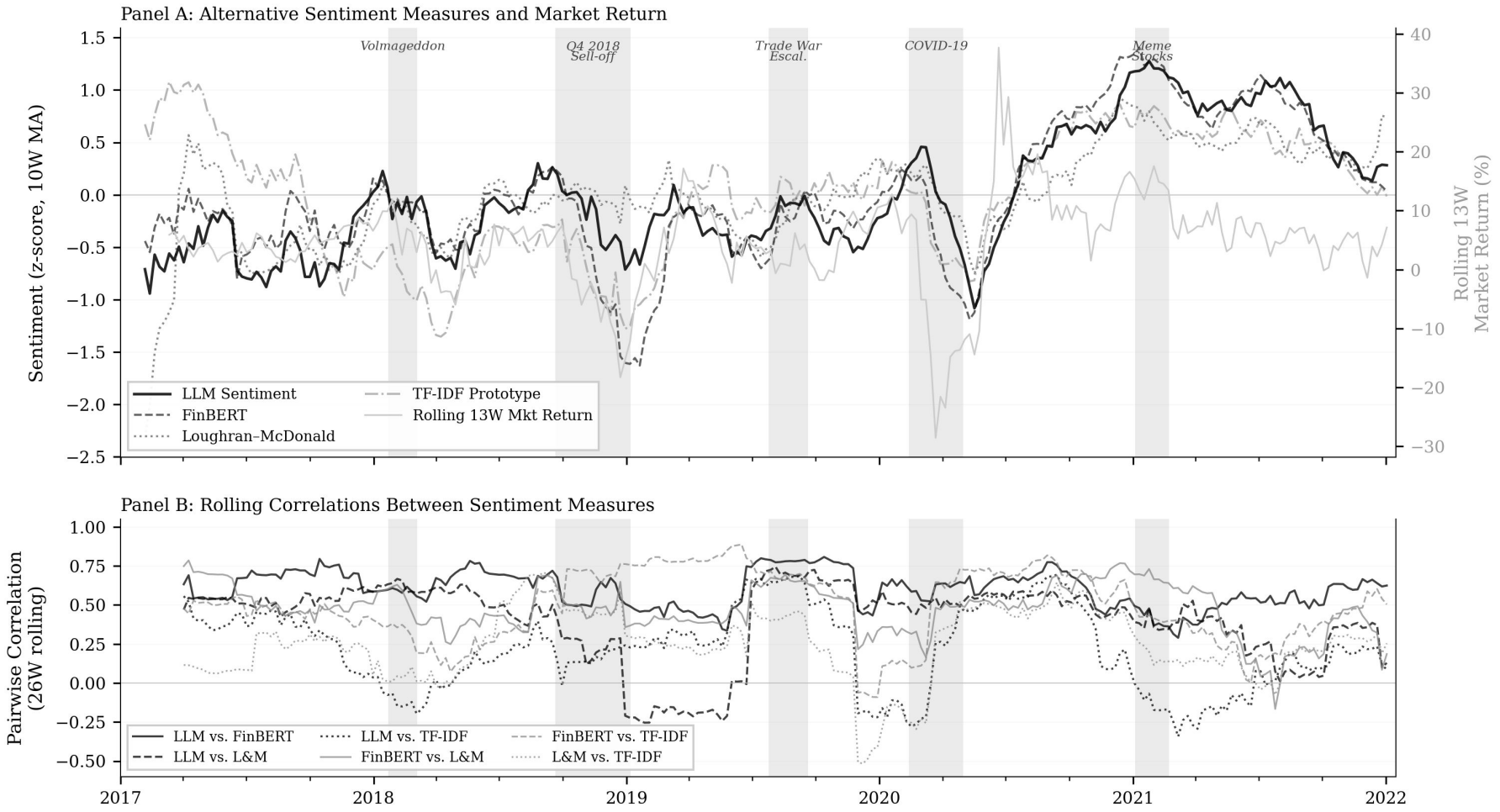


Figure OA.1: Alternative Sentiment Measures over Time

Notes: Panel A plots four sentiment measures over time, each standardized to zero mean and unit variance and smoothed with a 10-week moving average: the primary LLM sentiment score (solid), FinBERT (dashed), the Loughran–McDonald dictionary score (dotted), and the TF-IDF prototype sentiment (dash-dot). The light gray line shows the rolling 13-week compound market return (right axis). *Panel B* displays 26-week rolling pairwise correlations between the four sentiment measures. Gray bands mark five salient market events. The measures generally co-move positively, though correlations decline during high-volatility episodes such as the COVID-19 crash.

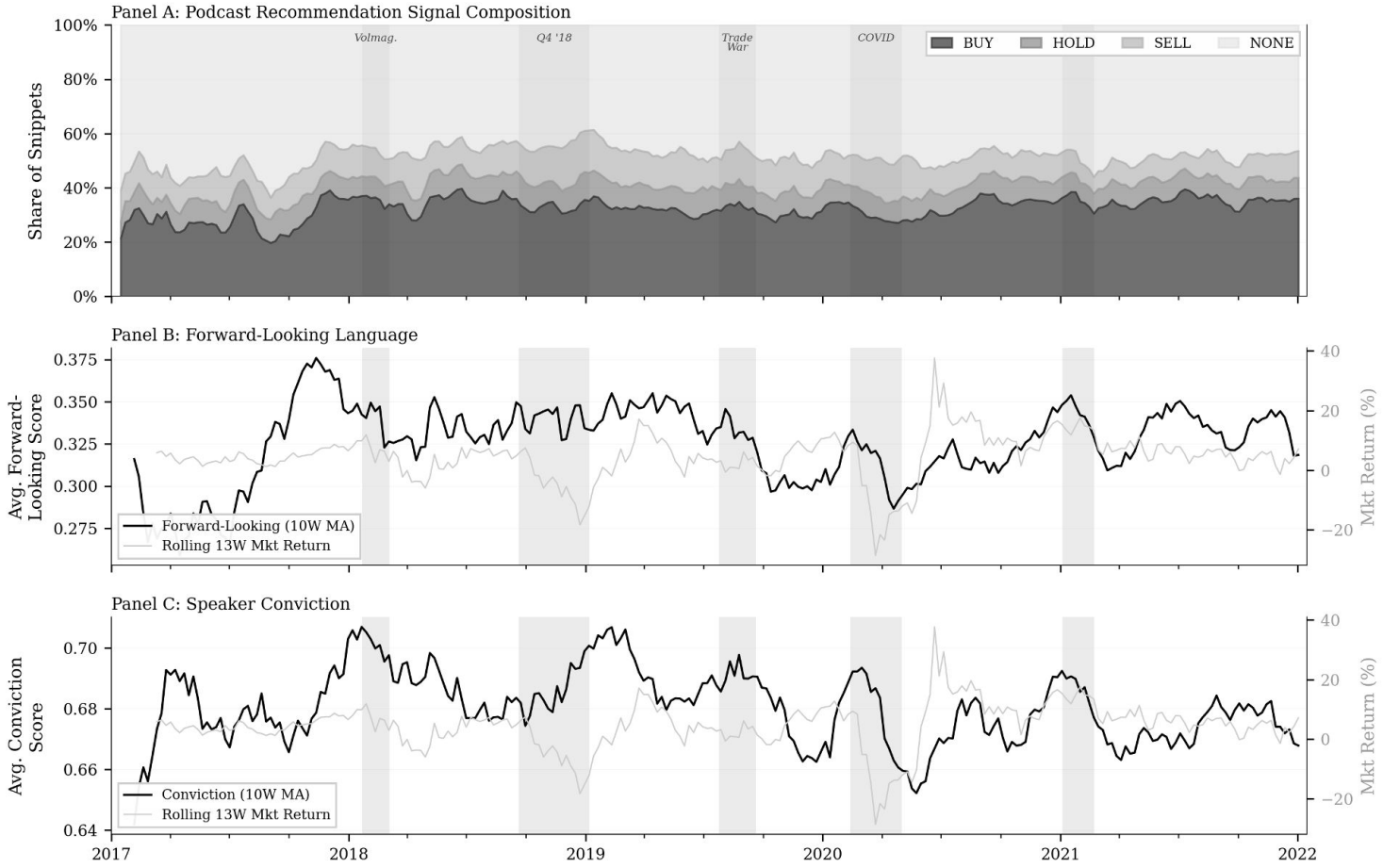


Figure OA.2: Podcast Content Dimensions over Time

Notes: Panel A shows the weekly composition of LLM-classified recommendation signals (BUY, HOLD, SELL, NONE) as stacked areas, smoothed with a 4-week rolling mean. The persistent dominance of BUY signals reflects the documented optimism bias (approximately 3:1 BUY-to-SELL ratio). *Panel B* plots the average forward-looking language score (10-week moving average, solid line) alongside the rolling 13-week market return (gray line, right axis). *Panel C* plots the average speaker conviction score (10-week moving average). Gray bands mark five salient market events. Forward-looking language increases during periods of market stress, consistent with hosts adopting a more predictive stance during uncertainty. Conviction is relatively stable over time with a slight upward trend.

Table OA.2: Descriptive Characteristics of the Investment Strategies in Section 4

Strategy	Type	Rebalancing	Trading Days	Pct. Days Invested	Avg. Stocks	Med. Stocks	Min Stocks	Max Stocks	Avg. Long	Avg. Short	Avg. Max Wt.
Buy & Hold (Market)	Long-only	None	1635	100	Market	Market	Market	Market	–	–	–
Podcast-Tilted Market (t+1)	Long-only (tilted)	Daily	1634	100	2189	2230	1929	2430	–	–	4.8%
Cumulative Belief (t+1)	Long-only	Monthly	1614	83	949	971	58	1650	–	–	–
Long-Short (t+1)	Long-short	Daily	1201	73	24	23	2	65	12	12	–
Contrarian Podcast (t+1)	Long-only	Monthly	1614	83	295	345	19	423	–	–	–

Notes: This table reports descriptive implementation characteristics of the portfolio strategies analyzed in Section 4. *Trading Days* is the number of sample trading days for which the strategy return is observed. *Pct. Days Invested* is the percentage of trading days on which the strategy holds at least one active position. *Avg. Stocks*, *Med. Stocks*, *Min Stocks*, and *Max Stocks* report the cross-day distribution of the number of stocks held in the portfolio. For the long–short strategy, *Avg. Long* and *Avg. Short* report the average number of stocks in the long and short legs. *Avg. Max Wt.* reports the average across days of the largest single-stock portfolio weight. For the buy-and-hold market benchmark, entries marked *Market* indicate that the strategy holds the value-weighted CRSP market portfolio throughout and that stock-level breadth statistics are therefore not separately tabulated.