The Value of Information from Sell-side Analysts

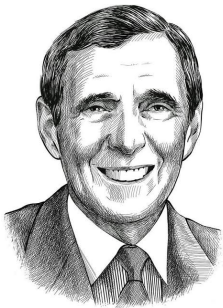
Linying Lyu[1]

[1]Washington University in St Louis

December 30, 2024

Background

- Sell-side analysts play an important role in modern capital markets.
  - Top-tier investment banks allocate more than $200 million each year to equity research.



"Wall Street analysts know their companies. You should cut a research report in two. The first part, the information about the company and its prospects, is probably pretty good. The second part, the recommendation, should be used as kindling. We use analyst information, but we don't use the recommendations very often."

– David Dreman

This Paper

- Research Questions:
  - What is the **strategic dollar value** of information from sell-side analysts?
    The **lower bound** of annualized profit of early access to analyst reports.
  - Do **written reports** provide incremental value beyond analyst forecasts?
    The **written report** is more valuable than forecast revisions.
  - What content in analyst reports is the most valuable?
    Analysts' interpretation of realized **income statements.**
- These questions are analyzed using LLMs and ML techniques.

This Paper

- Research Questions:
  - What is the **strategic dollar value** of information from sell-side analysts?
    The **lower bound** of annualized profit of early access to analyst reports.
  - Do **written reports** provide incremental value beyond analyst forecasts?
    The **written report** is more valuable than forecast revisions.
  - What content in analyst reports is the most valuable?
    Analysts' interpretation of realized **income statements.**
- These questions are analyzed using LLMs and ML techniques.

This Paper

- Research Questions:
  - What is the **strategic dollar value** of information from sell-side analysts?
    The **lower bound** of annualized profit of early access to analyst reports.
  - Do **written reports** provide incremental value beyond analyst forecasts?
    The **written report** is more valuable than forecast revisions.
  - What content in analyst reports is the most valuable?
    Analysts' interpretation of realized **income statements.**
- These questions are analyzed using LLMs and ML techniques.

This Paper

- Research Questions:
  - What is the **strategic dollar value** of information from sell-side analysts?
    The **lower bound** of annualized profit of early access to analyst reports.
  - Do **written reports** provide incremental value beyond analyst forecasts?
    The **written report** is more valuable than forecast revisions.
  - What content in analyst reports is the most valuable?
    Analysts' interpretation of realized **income statements.**
- These questions are analyzed using LLMs and ML techniques.

What is the total profit of early information acquisition?

- **Early access to analyst views**
- **Comparable insights from independent research**

Strategic Value of Information

- Back, Cao, and Willard (2000) model
  - A continuous-time version of Kyle (1985).
  - Informed traders possess **correlated partial information** on an asset.
- Kadan and Manela (2024) approach
  - A lower bound of aggregated expected profit at the stock level.

$$\Omega = \frac{\text{var}(\tilde{v})}{\lambda(0)} P(0)$$

  - The ratio of return volatility to price impact.

BCW Model

- Consider an asset with a normally distributed value $\tilde{v}$ to be revealed at t_1. Trading on the risky asset takes place continuously during the time period $[t_0, t_1)$.
- Three types of traders:
  - n risk-neutral **informed traders** who learn a signal correlated with $\tilde{v}$ at t_0.
  - **noise traders** submitting exogenous orders.
  - risk-neutral **market makers** who set prices to clear the market.

Informed Traders and Trading Strategy

- $n \geq 1$ risk-neutral informed traders, each receiving a signal $\tilde{s}^i$. Signals $\tilde{s}^i$ are joint-normally distributed with correlation $0 \leq \rho < 1$.
- Expected value $\tilde{v}$ of the asset:

$$\tilde{v} = \sum_{i=1}^{n} \tilde{s}^i.$$

- Trader i's strategy linear in price $P(t)$ and signal $\tilde{s}^i$:

$$\theta^i(t) = \alpha(t)P(t) + \beta(t)\tilde{s}^i.$$

- Profit maximization for trader i:

$$E\left[\int_0^1 (\tilde{v} - P(t))\left(\alpha(t)P(t) + \beta(t)\tilde{s}^i\right) dt\right].$$

Market Makers and Price Dynamics

- Risk-neutral market makers set the asset price $P(t)$ based on total order flow.
- The cumulative noise trading process $Z(t)$ follows a Wiener process:

$$dZ(t) = \sigma_z dW(t),$$

  where $W(t)$ is a Wiener process, and σ_z is the volatility of noise trading.
- The price $P(t)$ evolves according to the combined order flow:

$$dP(t) = \lambda(t) \left[dZ(t) + \sum_{i=1}^{N} \left(\alpha(t) P(t) + \beta(t) \tilde{s}^i \right) dt \right].$$

Equilibrium

- The equilibrium in this model is defined by two key conditions:
  1. $P(t) = V(t)$ for all t, where $V(t)$ is the conditional expectation of the asset's liquidation value $\tilde{v}$, given the information from the aggregate order process.
  2. Optimal trading strategy θ^i, given θ^{-i} and given λ.
- The equilibrium ex-ante expected value of information:

$$\Omega = c(n,\rho) \underbrace{\left[\frac{\text{var}(\tilde{v})}{\lambda(0)} P(0)\right]}_{\textbf{Lower Bound}},$$

where

$$c(n,\rho) = e^{\zeta} \int_1^{\infty} x^{-\frac{2}{n}} e^{-x\zeta} dx > \mathbf{0.92}.$$

Analyst Information Value Estimation

- Dollar Value Estimation:

$$\widehat{\Omega}_{it} = \frac{r_{it}^2 - \left(r_{it} - \frac{\sum_{j=1}^{N} \widehat{r}_{ijt}}{N} \right)^2}{\widehat{\lambda}_{it} / p_{it\_}},$$

where:

- $\widehat{\Omega}_{it}$: Dollar value of information for stock i on day t.
- r_{it}: Realized cumulative abnormal return ($CAR_{[-1,+1]}$).
- $\widehat{r}_{ijt}$: **Estimated $CAR_{[-1,+1]}$ from analyst report j on day t.**
- N: Number of analyst reports on stock i on day t.
- $\widehat{\lambda}_{it}$: **Estimated price impact.**
- $p_{it\_}$: Closing price of the day prior to the window $[t-1, t+1]$.

Price Impact Estimation

- Price Impact Regression:

$$r_{itk} = \widehat{\lambda}_{it} y_{itk} + \epsilon_{itk},$$

where:

- $r_{itk} = p_{it\tau_k} - p_{it\tau_{k-1}}$: 1-minute log return.
- $y_{itk} = Y_{it\tau_k} - Y_{it\tau_{k-1}}$: Change in cumulative signed order flows.
- $p_{it\tau_k}$: Log price of stock i within the window $[t-1, t+1]$ at time τ.
- $\widehat{\lambda}_{it}$: Estimated price impact.

Econometric Modeling

- Machine Learning Models
  - Baseline: Train Ridge model for the following equation each year with an expanding training window.

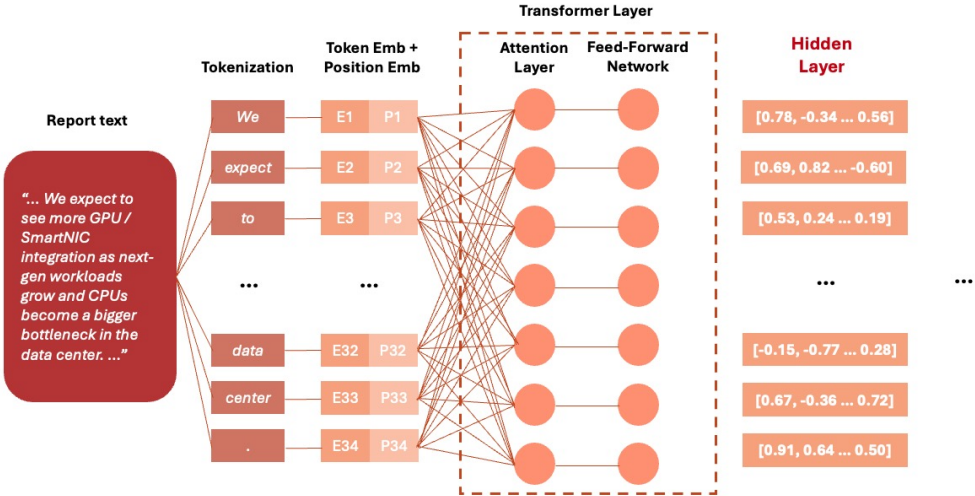
$$CAR_{[-1,+1],it} = \beta_0 + \beta' y_{ijt}^{\text{AI}} + \epsilon_{ijt},$$
$$\hat{\beta} = \underset{\beta}{\text{argmin}} \left\{ \left\| CAR_{[-1,+1],it} - \beta_0 - y_{ijt}^{\text{AI}} \beta \right\|_2^2 + \theta \|\beta\|_2^2 \right\}.$$

  - Using the 60% of data as the initial training sample. The OOS test started in 2015 and ends in 2023.

Analyst Infomation

- Report text measure: 5,120-dimension of text embeddings from LLaMA-2-13B model.
- Forecast Revision measures:
  - REC_REV: Recommendation revision, the current report's recommendation minus the last recommendation.
  - EF_REV: Earnings forecast revision, the current report's EPS forecast minus the prior EPS forecast, scaled by the stock price 50 days before the report release.
  - TP_REV: Target price revision, the current report's target price minus the last target price, scaled by the stock price 50 days before the report release.
- Other Numerical measures: $Boldness$, SR, $ERet$, $Prior\_CAR$, $Size$, $BtoM$, SUE ...

Text Representation



Data

- Two types of information from analysts
  - **Qualitative information**: 122,252 sell-side analyst reports from Investext on S&P 100 companies from 2000-01 to 2023-12. ∘ Summary Statistics
  - **Quantitative information**: analyst forecasts from I/B/E/S.
- Earnings information
  - Earnings conference call transcripts from Seeking Alpha.
- Stock returns and fundamentals from CRSP and Compustat.
- Intraday data from the NYSE TAQ database.

Summary Statistics for Analyst Information Values

- A stock typically has an average of 15 days with report releases each year.
- The lower bound investors would expect to profit annually for early access to analyst information on an S&P 100 stock is **$6.89 million**.

	Mean	SE	95% CI	99% CI
Information Value ($M)	**0.47**	0.05	[0.38, 0.56]	[0.35, 0.58]
Information Value of Text ($M)	**0.38**	0.04	[0.30, 0.46]	[0.28, 0.48]
Information Value of Revisions ($M)	**0.34**	0.04	[0.26, 0.43]	[0.23, 0.46]

- Delta method

The Value of Analyst Information in the Cross-section of Stocks

- Size effect:
  - Larger stocks typically have better liquidity.
  - Smaller stocks may have larger market reactions due to less publicly available information.
- Decomposition:

$$\log \widehat{\Omega}_{it} = \underbrace{\log \left[r_{it}^2 - \left(r_{it} - \frac{\sum_{j=1}^{N} \widehat{r}_{ijt}}{N} \right)^2 \right]}_{\text{log Explained Return Variance}} - \underbrace{\log \frac{\widehat{\lambda}_{it}}{P_{it\_}}}_{\text{log Price Impact}}$$

The Value of Analyst Information in the Cross-section of Stocks

- The value of analyst information is higher for large firms.
  - Smaller Price impact for large firms.

	log Info Value	log Return Variance	log Price Impact
	(1)	(2)	(3)
Size	**0.864*** **	0.044	**-0.826*** **
	(9.62)	(0.78)	**(-9.46)**
Book-to-market	0.457	0.180	-0.274
	(1.36)	(0.70)	(-1.33)
Year FE	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes
N	8030	8030	8030
Adjusted R^2	0.168	0.092	0.450

The Value of Analyst Information in the Cross-section of Analysts

- The value of analyst information is higher for bold analysts.
  - Higher explained return variance for bold analysts.

	log Info Value	log Return Variance	log Price Impact
	(1)	(2)	(3)
Bold	**0.290*** **	**0.231*** **	-0.059
	(4.090)	**(5.910)**	(-1.340)
Brokersize	0.001	-0.000	-0.002
	(0.880)	(-0.220)	(-1.680)
Year FE	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes
Analyst FE	Yes	Yes	Yes
N	13606	13587	12812
Adjusted R^2	0.201	0.122	0.442

The Value of Analyst Information around Earnings Announcements

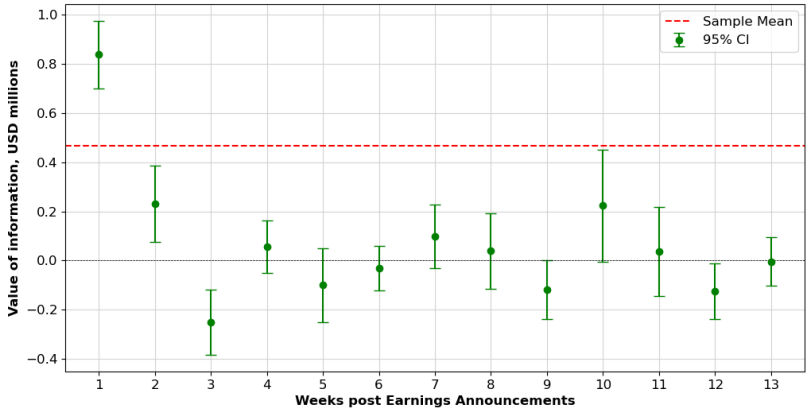


Figure: Dollar Value of Analyst Reports after Earnings Announcements

The Value of Information around Earnings Announcements

- Information value from analysts is higher following earnings announcements, especially when trading volumes are high.
  - Higher explained return variance

	log Info Value	log Return Variance	log Price Impact
Week	0.380***	0.473***	0.083**
	(3.540)	(5.380)	(2.000)
Trading Volume	-0.004	0.005	0.009**
	(-0.960)	(1.410)	(2.030)
Week × Trading Volume	0.017***	0.011**	-0.005**
	(2.950)	(2.410)	(-2.370)
Year FE	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes
N	8030	8030	8030
Adjusted R^2	0.158	0.115	0.396

Which type of information holds more value?

Quantitative or qualitative?

Out-of-Sample R-Squared

- Following Gu, Kelly, and Xiu (2020), define the OOS R^2 as:

$$R^2_{\text{oos}} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}} \left(CAR_{i,t} - \widehat{CAR}_{i,t}\right)^2}{\sum CAR^2_{i,t}},$$

  where $\mathcal{T}$ is the test set of (i,t) which has not been used in training and validation of the model.

- Diebold-Mariano test for pair-wise model comparison:

$$d_{12,t} = \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\left(\hat{e}^{(1)}_{i,t}\right)^2 - \left(\hat{e}^{(2)}_{i,t}\right)^2 \right)$$

Information Content of Numerical and Textual Information

- **Rev**: Revisions of EPS forecasts, stock recommendations, and target price
- **Num**: Rev + 14 additional numerical measures
- **Text**: Text embeddings of analyst reports

Year	**Rev only**	t-stat	**Num only**	t-stat	**Text only**	t-stat	**Rev + text**	t-stat	t-stat	t-stat	t-stat
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(5)-(1)	(7)-(1)	(7)-(5)
2015	10.31%	4.27	7.75%	2.59	12.63%	6.39	10.63%	2.98	3.63	0.16	-1.25
2016	14.61%	11.26	15.35%	11.26	11.98%	3.93	17.08%	9.67	-1.15	3.82	2.93
2017	8.99%	6.23	9.99%	6.41	11.11%	6.40	11.98%	7.09	4.99	5.96	2.38
2018	10.05%	3.28	10.52%	3.60	10.87%	5.85	13.64%	5.95	0.87	5.77	4.47
2019	9.94%	20.14	9.97%	14.76	12.16%	17.68	14.44%	26.17	2.68	19.77	4.83
2020	5.52%	3.88	6.10%	4.55	3.82%	5.18	6.34%	6.16	-1.75	1.57	7.11
2021	5.43%	5.83	5.96%	6.82	8.50%	6.21	11.94%	27.16	1.48	7.48	3.28
2022	9.78%	6.03	9.16%	4.97	14.88%	10.34	16.95%	10.09	9.21	8.01	5.30
2023	6.68%	5.48	7.19%	4.00	9.30%	4.17	10.76%	6.09	2.13	4.87	3.08
Overall	9.01%	9.45	9.08%	9.44	10.19%	8.20	12.28%	8.87	1.66	3.95	3.77

Information Content of Earnings Announcement Transcripts

- **Compare analyst reports with the latest earnings conference call transcripts.**

$$CAR_{[-1,+1],it} = \beta_0 + \beta' y_{ijt}^{\text{AI+EA}} + \epsilon_{ijt},$$

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \left\{ \left\| CAR_{[-1,+1],it} - \beta_0 - y_{ijt}^{\text{AI+EA}} \beta \right\|_2^2 + \theta \|\beta\|_2^2 \right\}.$$

- **Analyst reports add significant incremental value.**

	Transcripts	Reports	Reports + Transcripts	Diff
	(1)	(2)	(3)	(3) - (1)
R^2_{OOS}	4.20%	9.72%	11.96%	7.76%
t-stat	5.16	3.24	6.42	5.12

What insights in analyst reports create value?

Shapley Value Decomposition

- **Text Embeddings**:

$$y^{emb} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{e}_i,$$

  - e_i: the embedding of token i across all the layers
  - N: the total number of words in the report.

- **Topic-specific Embeddings**:

$$y^{emb} = \sum_{p=1}^{P} y_p^{emb} = \frac{1}{N}\sum_{p=1}^{P}\sum_{i_p=1}^{N_p} e_{i_p},$$

  - y_p^{emb}: the embedding for topic p
  - N_p: the number of tokens belonging to topic p
  - i_p: the index of words within topic p

Shapley Value Decomposition

- **Sentence-Segmented Text Embeddings**:

$$y^{emb} = \sum_{i=1}^{n} \frac{Token_i}{\sum_{i=1}^{n} Token_i} y_i^{emb},$$

  - $Token_i$: the number of tokens in i^{th} sentence
  - y_i^{emb}: the embeddings of the i^{th} sentence.

- **Shapley Value of Topic** p:

$$\varphi_p\left(R^2_{\text{oos}}\right) = \sum_{S \subseteq P \setminus \{p\}} \frac{|S|!\,(P - |S| - 1)!}{P!} \left[R^2_{\text{oos}}\left(y_s^{\text{emb}} + y_p^{\text{emb}}\right) - R^2_{\text{oos}}\left(y_s^{\text{emb}}\right)\right].$$

ChatGPT Prompting

- **First step: Get manageable and meaningful topics**

> **Prompt 1:** Please read the provided text file of sell-side analyst reports carefully. Identify high-level, mutually exclusive topics covered in these reports. Ensure that each sentence from the text file can be assigned to one of these topics. Here is the report content: {text}.

- **Second step: Categorize each sentence into a single topic**

> **Prompt 2:** Please read the following sentence from the sell-side analyst report of the company {firm} ({ticker}) carefully. Determine which category of information it belongs to among the following 17 categories: {categories}. Consider the context of the report.
>
> Output your response in JSON format.
>
> Here is the sentence from the analyst report: {sentence}.

17 Topics Discussed by Analysts

Executive Summary

Company Overview

Industry Analysis

Competitive Landscape

Income Statement Analysis

Balance Sheet Analysis

Cash Flow Analysis

Financial Ratio

17 Topics Discussed by Analysts

Business Segments

Growth Strategies

Risk Factors

Management and Governance

ESG Factors

Valuation

Investment Thesis

Appendices and Disclosures

Topic Distribution

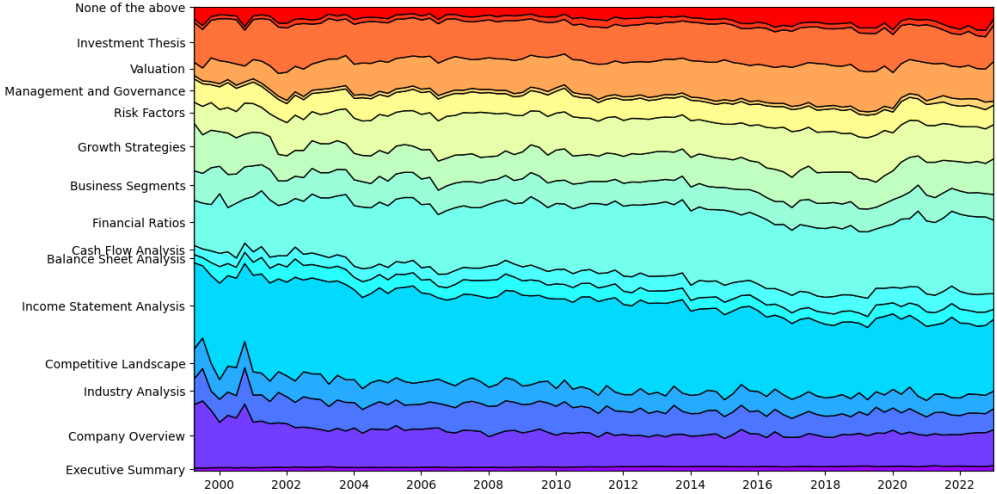


Figure: Analyst Discussion Across Topics

What content is valued by the market?

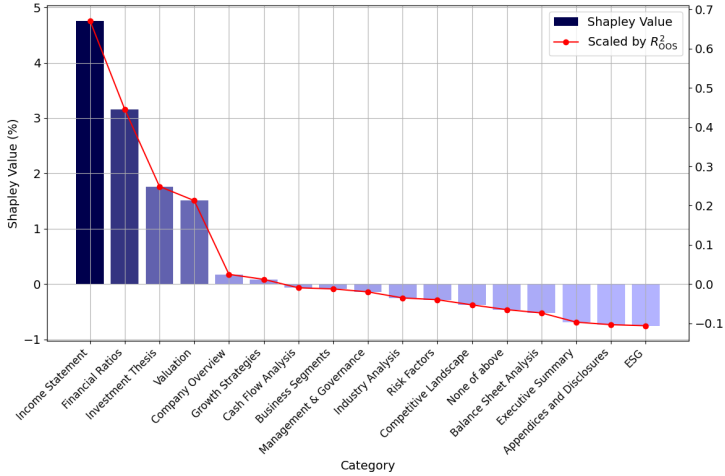


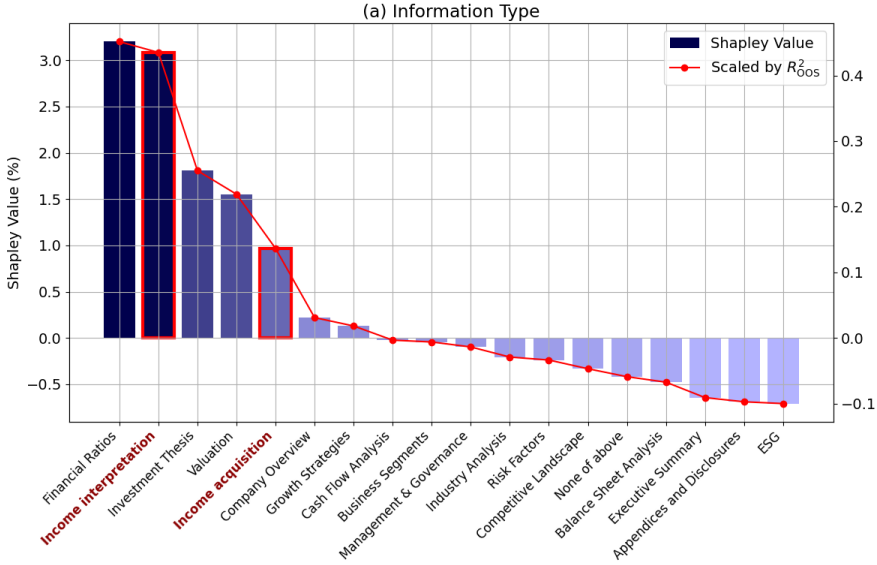
Figure: Shapley Values for Topic Importance

Interpretation and Discovery Roles

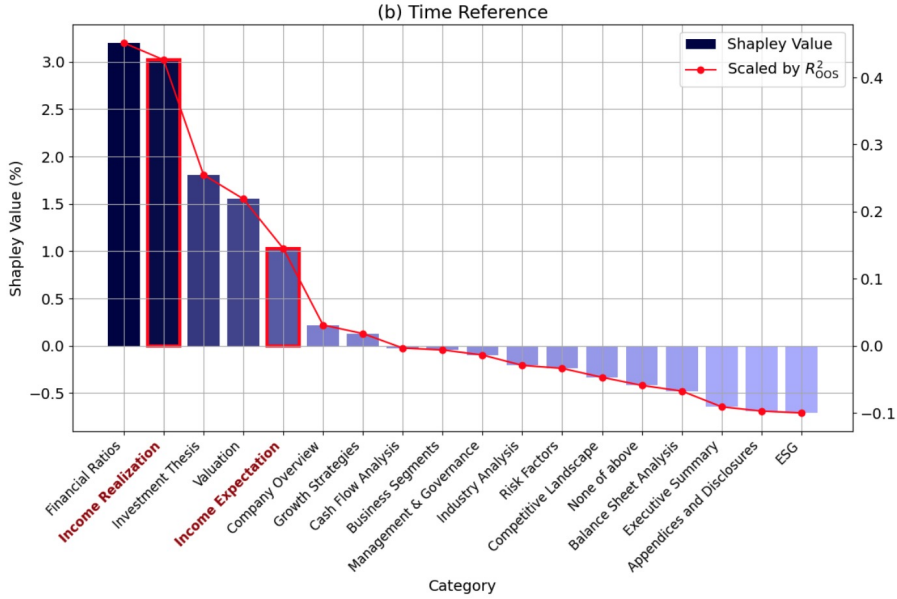
```
Prompt: Read the sentence below from a sell-side analyst report for {firm}
({ticker}) and classify it based on two criteria:
```

- **Information Type**
  - Information Acquisition: Sentences that report quantitative financial data
    e.g. 'Reported EPS of $1.40 beat consensus of $0.94 and our estimate of $1.17.'
  - Information Interpretation: Sentences that analyze or interpret financial data, such as trends or market impacts.
    e.g. 'Biogen reported 1Q12 MS franchise sales below our estimate and the Street, as Avonex sales were negatively impacted by unfavorable distribution channel dynamics.'
- **Time Reference**
  - Realized Income: Sentences that refer to historical results
    e.g. 'The company's revenue for the fiscal year 2022 was $500 million.'
  - Expected Income: Sentences that express future predictions
    e.g. 'The company expects to achieve revenue of $550 million in fiscal year 2023 based on current market conditions.'

Shapley Values for Sub-Topic Importance



Shapley Values for Sub-Topic Importance



Conclusion

- I quantified the dollar value of analyst information.
  - The annual expected profit for all strategic investors from analyst reports on an average S&P 100 constituent stock is larger than $ 6.89 million.
- Applying state-of-the-art LLMs to a large corpus of analyst reports:
  - I find significant explanatory power of text for contemporaneous stock returns OOS beyond and above quantitative forecasts.
- I propose a decomposition method and the findings emphasize the value of income statement analysis in written reports.

Thank you!

I welcome any questions, comments, or further discussion.

Summary Statistics

○ Back

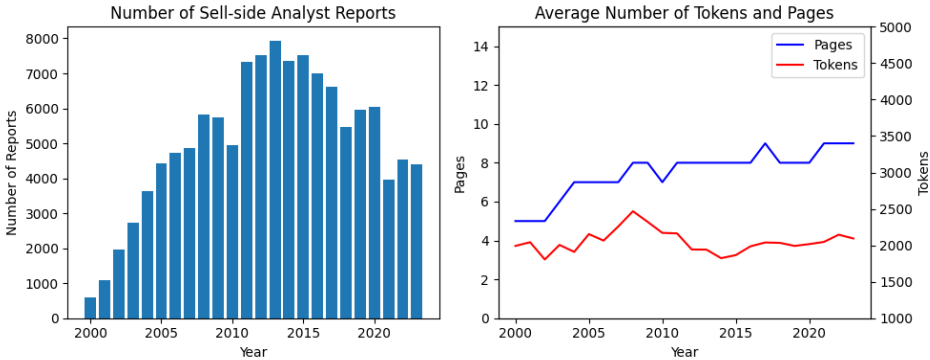


Figure: Summary Statistics for Report by Year (2000-2023)

Delta Method

◦ Back

- Direct computation of the mean information value Ω can be biased and unreliable due to the nature of the ratio of two random variables measured with error.
- Delta method: a statistical technique used to approximate the variance (or uncertainty) of a function of one or more random variables.
- Mean Information Value:

$$\mathbb{E}\widehat{\Omega}s = \frac{\mu_{vs}}{\mu_{\lambda_s}}$$

- Variance of Information Value:

$$\operatorname{Var}\widehat{\Omega}s = \frac{1}{\mu_{\lambda_s}^2}\left(\Sigma_{\lambda_s}\frac{\mu_{vs}^2}{\mu_{\lambda_s}^2} + \Sigma_{vs} - 2\Sigma_{v\lambda_s}\frac{\mu_{vs}}{\mu_{\lambda_s}}\right)$$