The use and usefulness of big data in finance: Evidence from financial analysts

Feng Chi, Byoung-Hyoun Hwang, and Yaping Zheng

2020.11 Working Paper

Presenter: YueYang

2021.7.22

Contents

- 1.Introduction
  - 1.1Background
  - 1.2Literatures
  - 1.3Motivations
  - 1.4Contributions
- 2.Data
- 3.Methods
- 4.Results
- 5.Conclusions

1.1Background

- With the advent of modern information technologies and recent advances in data analytics, we can increasingly track individuals' and businesses' activities through the digital footprints they leave behind.
- "Big data" or "alternative data" are being used in the financial sector.
- New platforms enabled by recent technologies have begun to successfully encroach on mainstream Wall Street's territories

1.2Literatures

- Alternative data contain unique and important information(Chiu, Teoh, Zhang, and Huang, 2018; Huang, 2018)
- Anecdotal accounts in the financial press routinely point to the transformative nature of alternative data (Dannemiller and Kataria, 2017; Ram and Wigglesworth, 2017)
- Some analysts have better access to management than others (e.g. Cohen, Frazzini, and Malloy 2010; Brown, Call, Clement, and Sharp 2015)
- Alternative data are of limited use given the noise and biases in the data (Hope, 2016).

1.2Literatures

- New platforms enabled by recent technologies have begun to successfully encroach on mainstream Wall Street's territories:
  - lending (Morse, 2015)
  - financial advice (D'Acunto and Rossi, 2020)
  - investment advice (Pacelli, 2020)

1.3Motivations

- Q1: How widely used alternative data are by financial firms?
- Q2: What analyst and firm characteristics determine the use of alternative data?
- Q3: The use of alternative data and forecast accuracy?
- Q4: The use of alternative data and stock market reactions?

1.4Contributions

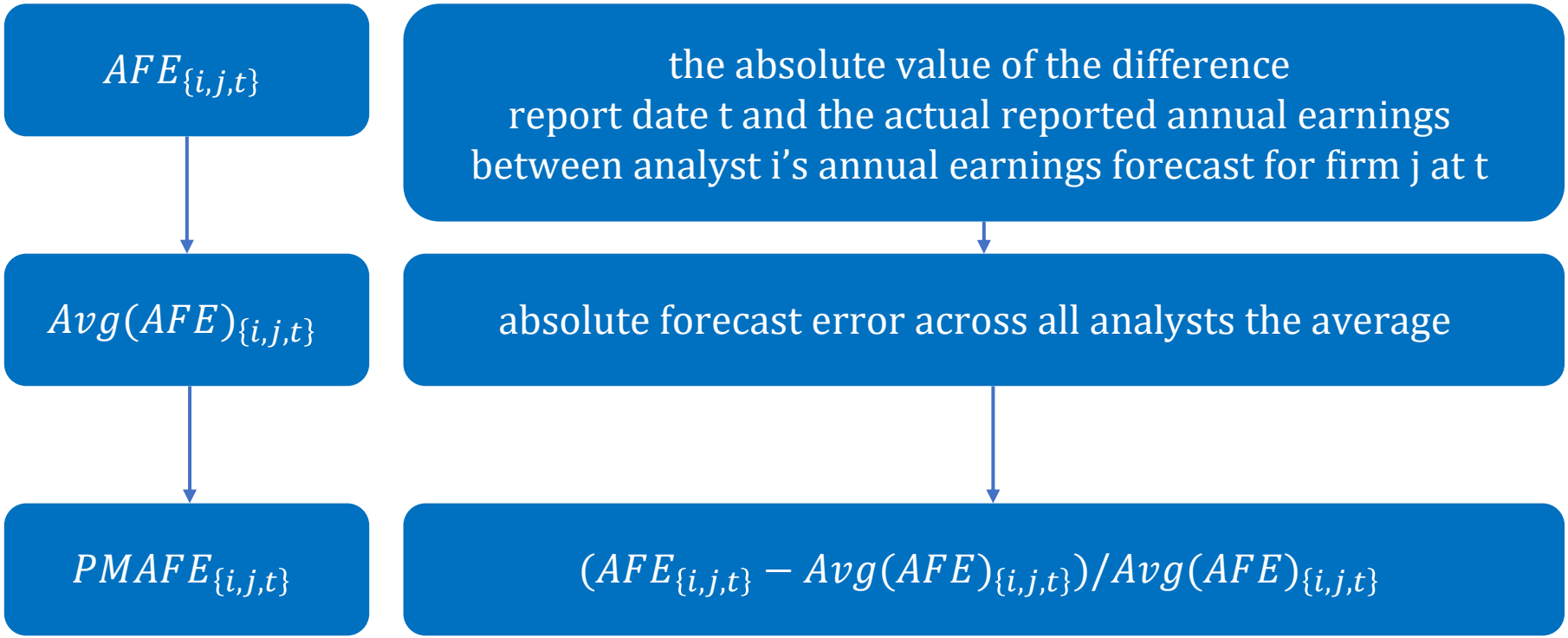
- Our study is the first systematic analysis on the actual use of alternative data.
- Capture performance implications relatively cleanly by studying how the use of alternative data relates to earnings forecast accuracy.
- Our evidence can help guide discussions on the impact of big data on investor welfare and market quality.

2.Data

- Constituents of the Dow Jones Industrial Average index (DJI)
- For each report: the report date, the report title, the analyst names and their work phone numbers, the name of the broker, and the full text.
- Merge our “reports data” with annual earnings forecast data from the IBES database, financial market data from CRSP and financial statement data from Compustat
- Our final sample comprises 64,036 written reports compiled by 1,002 distinct analysts working for 55 distinct brokers.
- 2009.06-2019.05.

3.Methods

- Variables:
- Y: an analyst's earnings forecast accuracy

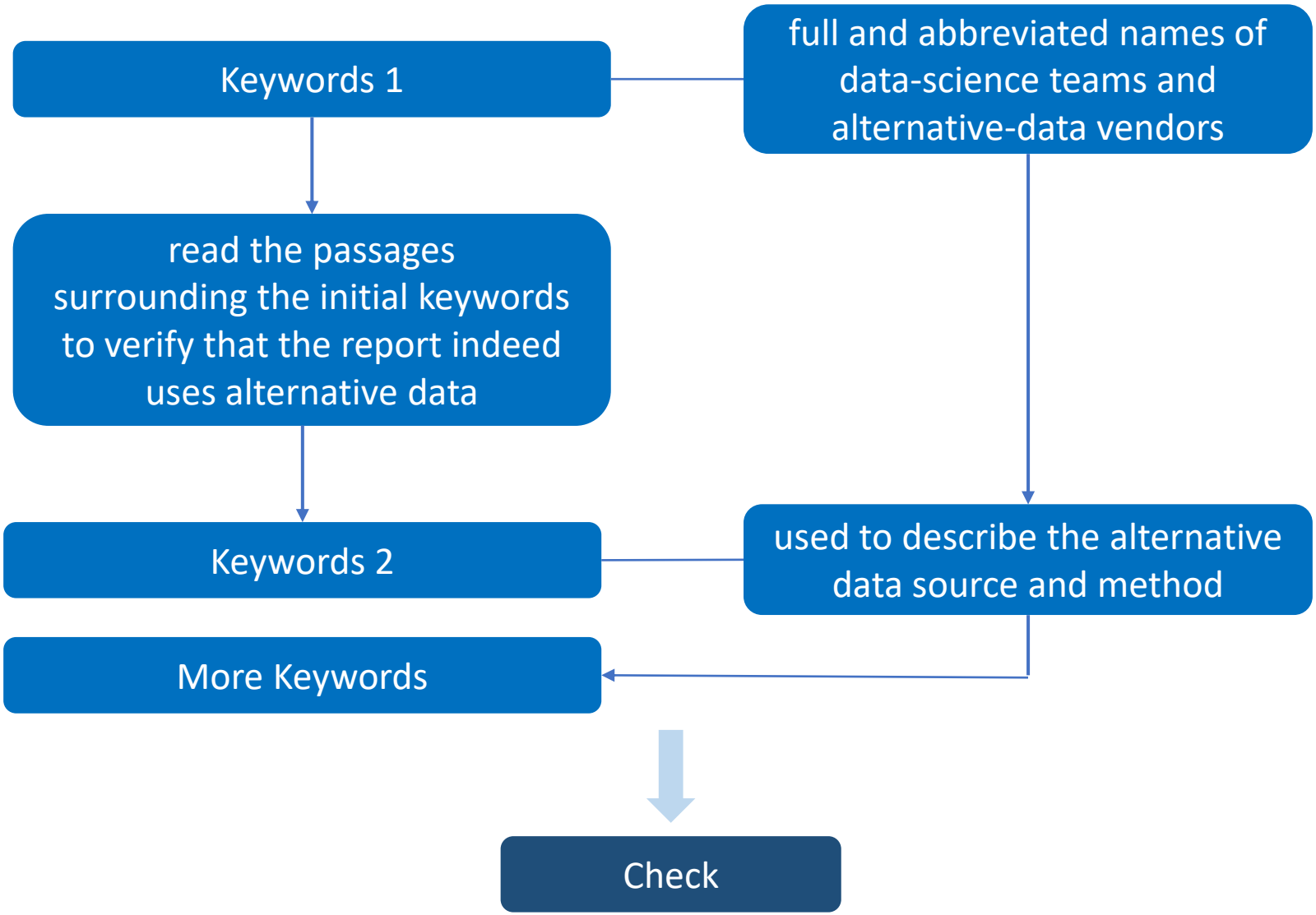


- $Accuracy_{\{i,j,t\}} = PMAFE_{\{i,j,t\}} * (-1)$

3.Methods

- Variables:
- X: an analyst's use of alternative data
- two types of keywords to identify the use of alternative data: (1) names of in-house data-science teams and names of "external" alternative-data vendors
- (2) keywords that describe a data source or a method associated with alternative data (e.g. "web scraping," "Facebook likes").
- $I\left(Alternative\ Data_{i,j,t}\right)$

- **How to judge whether an analyst use alternative data in his report?**



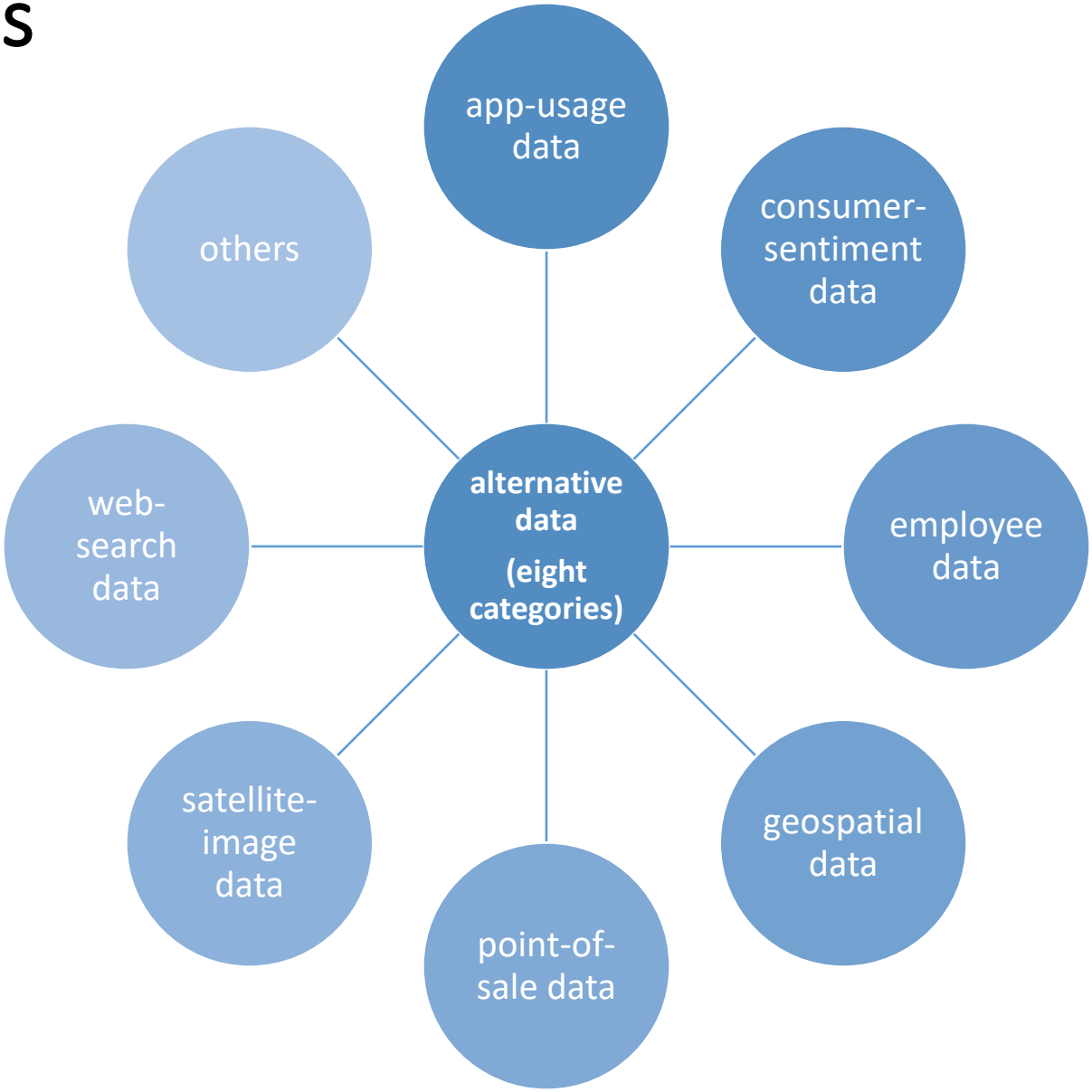
4.Results

Q1: How widely used alternative data are by financial firms?

	Number of Forecasts …		Fraction of Forecasts…
	… Supported by Alternative Data	… Not Supported by Alternative Data	… Supported by Alternative Data
Panel A: By Year			
2009/2010	515	7,634	6%
2011	615	6,239	9%
2012	488	6,769	7%
2013	490	6,348	7%
2014	497	6,058	8%
2015	694	5,998	10%
2016	729	5,691	11%
2017	659	5,444	11%
2018/2019	952	8,216	10%
Panel B: By Industry Sector			
Energy	40	2,512	2%
Materials	29	2,614	1%
Industrials	580	8,398	6%
Consumer Discretionary	443	4,194	10%
Consumer Staples	841	7,199	10%
Health Care	661	7,487	8%
Financials	103	8,082	1%
Information Technology	2,513	13,597	16%
Communication Services	429	4,314	9%
∑	5,639	58,397	9%

- The ratio of using alternative data in analyst reports is basically showing an increasing trend.
- The information technology department use most.
- Energy (2%), financial (1%) and materials (1%) companies rarely use alternative data.

4.Results


app-usage data
consumer-sentiment data
employee data
geospatial data
point-of-sale data
satellite-image data
web-search data
others
alternative data (eight categories)

4.Results

Q1: How widely used alternative data are by financial firms?

- web-search data(34%)
- others (23%)
- employee (10%)
- app usage (8%)
- geospatial (5%)
- satellite image (3%)

Alternative Data Category	Number of Forecasts Supported by Alternative Data	Percentage
App Usage	476	8%
Consumer Sentiment	1,062	19%
Employee	543	10%
Geospatial	257	5%
Point of Sale	1,080	19%
Satellite Image	171	3%
Web Search	1,944	34%
Others	1,322	23%

4.Results

- **Q2: What analyst and firm characteristics determine the use of alternative data?**
- $I(Alternative\ Data_{i,j,t}) = \alpha + \beta Analyst\ Char_{i,j,t} + \gamma Firm\ Char_{i,j,t} + \varepsilon_{i,j,t}$

Analyst Char	Firm Char
Analyst/Firm Experience	Size
Analyst Experience	M/B
Firms Covered	Momentum
Forecast Frequency	Analyst Following
Broker Size	Earnings Volatility
I(In-House Data Science Team)	
∑ Local Colleagues using Alternative Data	
∑ Non-Local Colleagues using Alternative Data	

4.Results

- The use of alternative data by local colleagues strongly increases the likelihood that analysts adopt alternative data themselves

	All Analysts	Analysts employed by broker with in-house data science team	Analysts employed by broker without in-house data science team
	(1)	(2)	(3)
Analyst/Firm Experience	-0.006	-0.000	-0.009
	(-0.73)	(-0.02)	(-1.22)
Analyst Experience	0.001	0.006	-0.002
	(0.17)	(0.66)	(-0.39)
#Firms Covered	0.132	-0.038	0.141
	(1.08)	(-0.14)	(1.15)
Forecast Frequency	-0.046	-0.008	-0.037
	(-0.81)	(-0.06)	(-0.60)
Broker Size	0.024	-0.363	0.046
	(0.37)	(-1.31)	(0.69)
I(In-House Data Science Team)	0.460***		
	(3.73)		
∑ *Local Colleagues using Alternative Data*	0.047***	0.063***	0.049**
	(2.75)	(2.58)	(2.27)
∑ *Non-Local Colleagues using Alternative Data*	0.038*	0.053*	0.011
	(1.79)	(1.73)	(0.32)
Size	0.075	0.082	0.088
	(1.28)	(0.62)	(1.44)
M/B	0.013	0.003	0.015
	(1.30)	(0.33)	(1.37)
Momentum	0.276*	-0.190	0.376**
	(1.87)	(-0.56)	(2.29)
# Analysts Following	0.626***	1.157***	0.491***
	(3.50)	(3.52)	(2.66)
Earnings Volatility	0.034***	0.084***	0.012
	(2.97)	(4.74)	(1.05)
Analyst-Firm Fixed Effects	No	No	No
Firm-Year Fixed Effects	No	No	No
N	64,036	9,078	54,958
Pseudo R^2	0.106	0.241	0.045

- Analysts more frequently turn to a typical information sources when they face greater competition.
- Analysts more frequently draw from alternative data when earnings are harder to forecast.

4.Results

- **Q3: The use of alternative data and forecast accuracy**
- $Accuracy_{i,j,t} = \eta_{i,j} + \theta_{j,t} + \beta I(Alternative\ Data_{i,j,t}) + \gamma\ Controls + \varepsilon_{i,j,t}$
- analyst-firm ("group"), $\eta_{i,j}$
- firm-year ("period"), $\theta_{\mathrm{i,j}}$
- Controls——analyst characteristics:
  - Forecast Age, Analyst/Firm Experience, Analyst Experience, #Firms Covered, Forecast Frequency, and Broker Size.
- What does $I(Alternative\ Data_{i,j,t})$ mean?
  - how much more accurate an analyst becomes in the postadoption period relative to the pre-adoption period compared with peer analysts covering the same firm over the same period.

	(1)	(2)	(3)	(4)
① *I(Alternative Data)*	0.214***			
	(3.57)			
I(Alternative Data Category = App Usage)		0.308***		
		(4.80)		
I(Alternative Data Category = Consumer Sentiment)		0.221*		
		(2.20)		
I(Alternative Data Category = Employee)		0.206*		
		(1.77)		
I(Alternative Data Category = Geospatial)		-0.014		
		(-0.20)		
I(Alternative Data Category = Point of Sale)		0.181***		
		(3.09)		
I(Alternative Data Category = Satellite Image)		0.034		
		(0.41)		
② *I(Alternative Data Category = Web Search)*		0.147**		
		(2.55)		
I(Alternative Data Category = Others)		0.184***		
		(5.73)		
③ ∑ *Alternative Data Categories*			0.176***	
			(2.86)	
I(Alternative Data Source = In-House Data Science Team)				0.237***
				(3.97)
④ *I(Alternative Data Source = Data Vendor)*				0.184***
				(3.46)
I(Alternative Data Source = Unknown)				0.199***
				(2.61)
Forecast Age	-0.246***	-0.245***	-0.183***	-0.245***
	(-10.75)	(-11.03)	(-4.02)	(-10.86)
① *Analyst/Firm Experience*	0.058***	0.058***	-0.025	0.058***
	(4.61)	(4.59)	(-0.81)	(4.61)
Analyst Experience	0.063	0.062	0.383	0.061
	(1.31)	(1.29)	(1.29)	(1.28)
#Firms Covered	0.038	0.039	-0.274	0.039
	(0.60)	(0.62)	(-1.42)	(0.62)
Forecast Frequency	0.032	0.031	-0.095	0.030
	(1.06)	(1.02)	(-1.18)	(0.98)
Broker Size	-0.082**	-0.084**	0.316	-0.080**
	(-2.13)	(-2.17)	(1.41)	(-2.14)
Analyst-Firm Fixed Effects	Yes	Yes	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes	Yes	Yes
N	64,036	64,036	5,639	64,036
Adjusted R^2	0.231	0.232	0.371	0.232

Six of eight

the number of distinct alternative data categories an analyst draws from in her report

alternative data resource

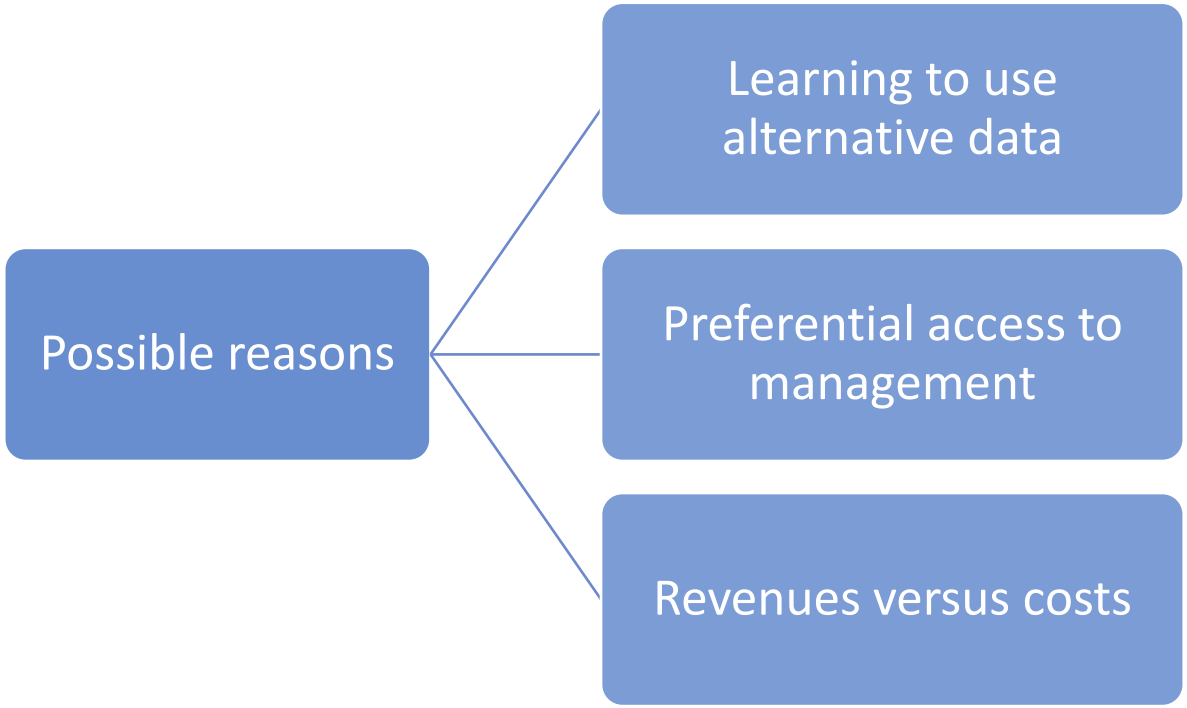
4.Results

- **Q3: The use of alternative data and forecast accuracy**
- Forecast accuracy increases significantly with the number of years an analyst has been covering a particular firm
- The use of alternative data from six of the eight categories are associated with statistically significant improvements in performance.
- Performance indeed improves further as analysts draw from a greater variety of alternative data.
- Adopting alternative data internally and adopting alternative data externally are accompanied with similar performance improvements.

4.Results

Q3: The use of alternative data and forecast accuracy

- Is the use of alternative data to improve forecast accuracy accidental or is there a real reason?



4.Results

H1: Learning to use alternative data

Learn from himself

Learn from others

	Learning from One's Own Experience (1)	Learning from Colleagues (2)
∑ *Years using Alternative Data*	0.222***	
	(2.91)	
∑ *Local Colleagues using Alternative Data*		0.064***
		(3.73)
∑ *Non-Local Colleagues using Alternative Data*		-0.043
		(-1.57)
Forecast Age	-0.174***	-0.199***
	(-4.14)	(-3.85)
Analyst/Firm Experience	-0.018	-0.037
	(-0.70)	(-1.38)
Analyst Experience	0.356	0.397
	(1.19)	(1.29)
#Firms Covered	-0.268	-0.318
	(-1.48)	(-1.58)
Forecast Frequency	-0.099	-0.099*
	(-1.42)	(-1.73)
Broker Size	0.288	0.359
	(1.45)	(1.41)
Analyst-Firm Fixed Effects	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes
N	5,639	5,639
Adjusted R^2	0.370	0.366

- Once analysts adopt alternative data, the positive effects of the use of alternative data continue to grow over time.
- While there is substantial learning across colleagues residing in the same locale, there is limited learning across colleagues residing in different locales

4.Results

H2: Preferential access to management

$$Accuracy_{i,j,t} = \theta_{j,t} + \beta_1 I(Alternative\ Data\ Within\ Executive\ Purview_{i,j,t}) + \beta_2 I(Alternative\ Data\ Outside\ Executive\ Purview_{i,j,t}) + \gamma\ Controls + \varepsilon_{i,j,t}$$

	With Preferential Access (1)	Without Preferential Access (2)	*F*-Test of Equality in Coefficient Estimate
I(Alternative Data Within Executive Purview)	0.059	0.163***	4.75**
	(1.01)	(4.48)	
I(Alternative Data Outside Executive Purview)	0.146**	0.117**	0.08
	(2.47)	(2.06)	
Analyst-Firm Fixed Effects	No	No	
Firm-Year Fixed Effects	Yes	Yes	
N	21,641	42,395	
Adjusted R^2	0.187	0.165	

app usage,
consumer sentiment, employee(manager sentiment and job postings),
point of sale,
satellite image,
web search

employee (employee sentiment),
geospatial or others

- For analysts with preferential access, data that fall within executive purview do not improve performance, but data that fall outside executive purview do improve performance.
- For analysts who lack preferential access, both data within executive purview and data beyond executive purview strongly improve performance.

4.Results

H3: Revenues versus costs

	Revenue Forecast Accuracy (1)	Cost Forecast Accuracy (2)	*F*-Test of Equality in Coefficient Estimate
I(Alternative Data)	0.148**	0.107	7.68***
	(2.15)	(1.49)	
Forecast Age	-0.119***	-0.107***	
	(-4.55)	(-4.85)	
Analyst/Firm Experience	0.032	0.055	
	(0.29)	(0.72)	
Analyst Experience	0.979***	0.756***	
	(4.99)	(4.70)	
#Firms Covered	-0.070	-0.024	
	(-0.73)	(-0.34)	
Forecast Frequency	0.076**	0.020	
	(2.12)	(0.65)	
Broker Size	-0.014	-0.047	
	(-0.22)	(-0.73)	
Analyst-Firm Fixed Effects	Yes	Yes	
Firm-Year Fixed Effects	Yes	Yes	
N	27,661	27,661	
Adjusted R^2	0.336	0.391	

- The adoption of alternative data comes with significantly more accurate revenue forecasts;
- The adoption of alternative data is not associated with more accurate cost forecasts

4.Results

- **Q4: The use of alternative data and stock market reactions**

$$CAR_{i,j,t} = \eta_{i,j} + \theta_{j,t} + \beta_1 I(Alternative\ Data_{i,j,t}) + \beta_2 \Delta_{i,j,t} + \beta_3 I(Alternative\ Data_{i,j,t}) \times \Delta_{i,j,t} + \gamma\ Controls + \varepsilon_{i,j,t}$$

- $CAR_{i,j,t}$: the cumulative market-adjusted return over days [0,+1]
  - Day 0: the date the earnings forecast, the price target, or the recommendation by analyst i for firm j is reported
- $\Delta_{i,j,t}$: either the percentage change in the earnings forecast, the percentage change in the target price, or the change in the "recommendation score."

4.Results

	Earnings Forecast Change		Target Price Change		Recommendation Change	
	(1)	(2)	(3)	(4)	(5)	(6)
I(Alternative Data) × *Δ*	7.777**	7.614**	3.227***	2.569**	0.667**	0.603*
	(2.47)	(2.47)	(3.10)	(2.59)	(2.16)	(1.96)
Δ	3.777**	4.231**	2.257***	2.898***	0.712***	0.715***
	(2.60)	(2.70)	(4.29)	(4.43)	(7.45)	(7.50)
I(Alternative Data)	0.104*	0.105**	0.096*	0.071	0.112**	0.104
	(1.89)	(2.12)	(1.75)	(1.27)	(2.07)	(1.66)
Forecast Age	-0.029	-0.015	-0.029	-0.020	-0.029	-0.016
	(-1.55)	(-0.82)	(-1.57)	(-1.02)	(-1.54)	(-0.88)
Analyst/Firm Experience	-0.000	-0.021*	-0.000	-0.023**	0.000	-0.022
	(-0.06)	(-1.79)	(-0.19)	(-2.21)	(0.22)	(-1.69)
Analyst Experience	0.001	0.019	0.001	-0.001	0.001	0.026
	(0.49)	(0.58)	(0.92)	(-0.02)	(0.80)	(1.01)
#Firms Covered	-0.061*	-0.093	-0.057	-0.072	-0.042	-0.085
	(-1.99)	(-1.21)	(-1.37)	(-0.79)	(-1.36)	(-0.96)
Forecast Frequency	0.021	0.095*	0.026	0.055	0.014	0.093
	(1.09)	(1.82)	(0.97)	(0.94)	(0.76)	(1.61)
Broker Size	0.020	-0.148*	0.024	-0.131*	0.026*	-0.131**
	(1.36)	(-1.96)	(1.46)	(-2.00)	(1.77)	(-2.12)
Firm Characteristics Controls	Yes	No	Yes	No	Yes	No
Analyst-Firm Fixed Effects	No	Yes	No	Yes	No	Yes
Firm-Year Fixed Effects	No	Yes	No	Yes	No	Yes
N	37,955	37,955	34,697	34,697	37,848	37,848
Adjusted R^2	0.023	0.045	0.024	0.046	0.024	0.044

- The market perceives analyst research as more informative when an analyst incorporates alternative data.

5.Conclusions

- Alternative data are widely used and that the adoption of alternative data is accompanied with strong improvements in forecast accuracy and heightened stock-market reactions.
- Our results suggest that analysts make insights gleaned from alternative data accessible to broad sections of the investor population.