Big Data and Financial Forecasting

Thierry Foucault

HEC Paris

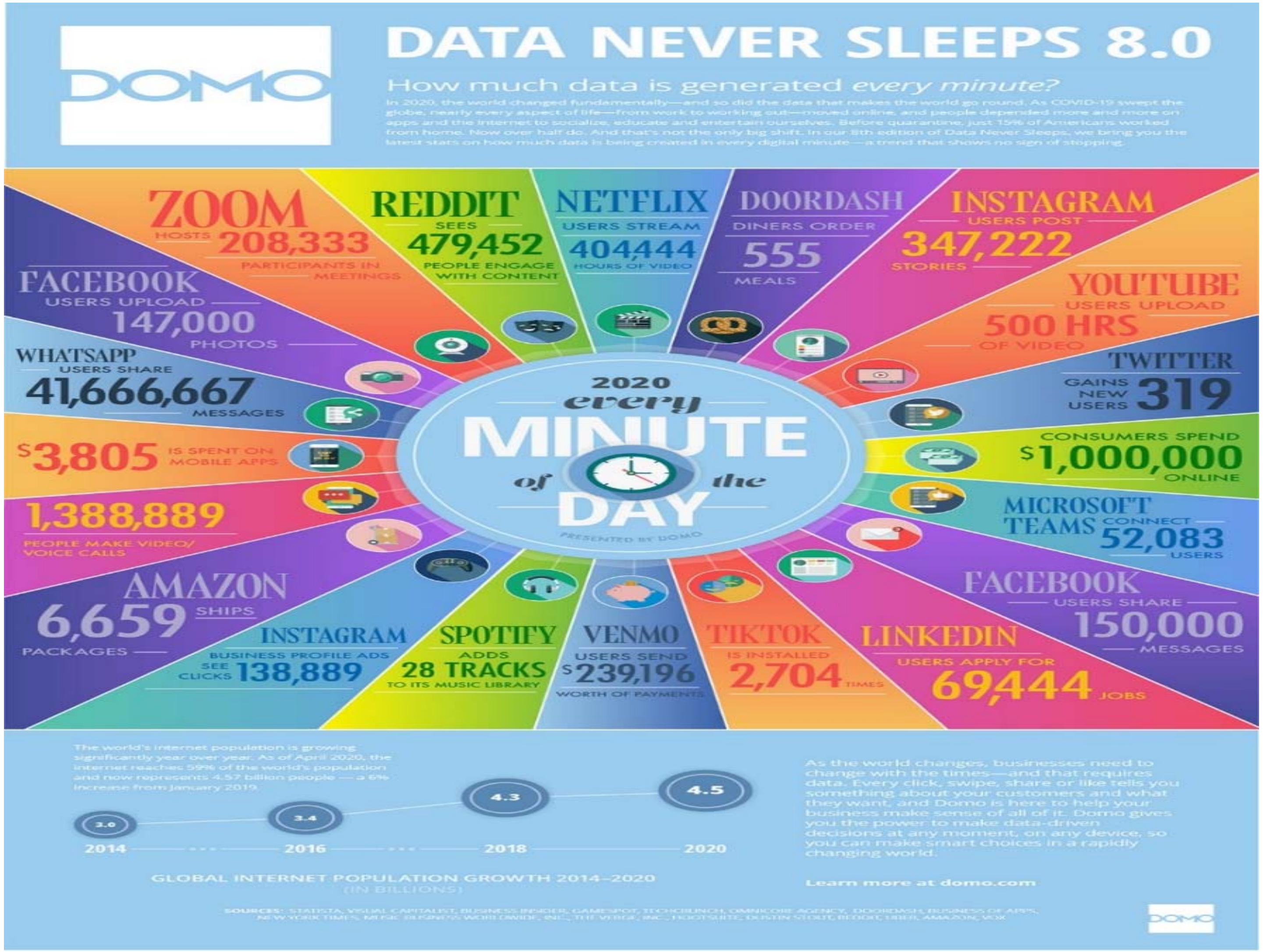
e.mail: foucault@hec.fr

Part of the research discussed in this presentation has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No 101018214)

Financial Information

- **Information plays a crucial in finance**
  - Required to forecast future cash-flows and price securities.
- **Markets for financial information**:
  - **Sellers** ("information intermediaries"): Securities analysts, credit rating agencies, data vendors (Bloomberg, Refinitiv etc.), trading platforms (Primary markets, MTFs, …) etc.
  - **Buyers:** Institutional investors, prop trading firms, brokers, banks etc.
- **Markets for financial information experience major changes due to the emergence of**
  - New data ("Alternative data")
  - New ways to capture and store data ("digitization")
  - New ways to process data to generate information (AI)

Big Data

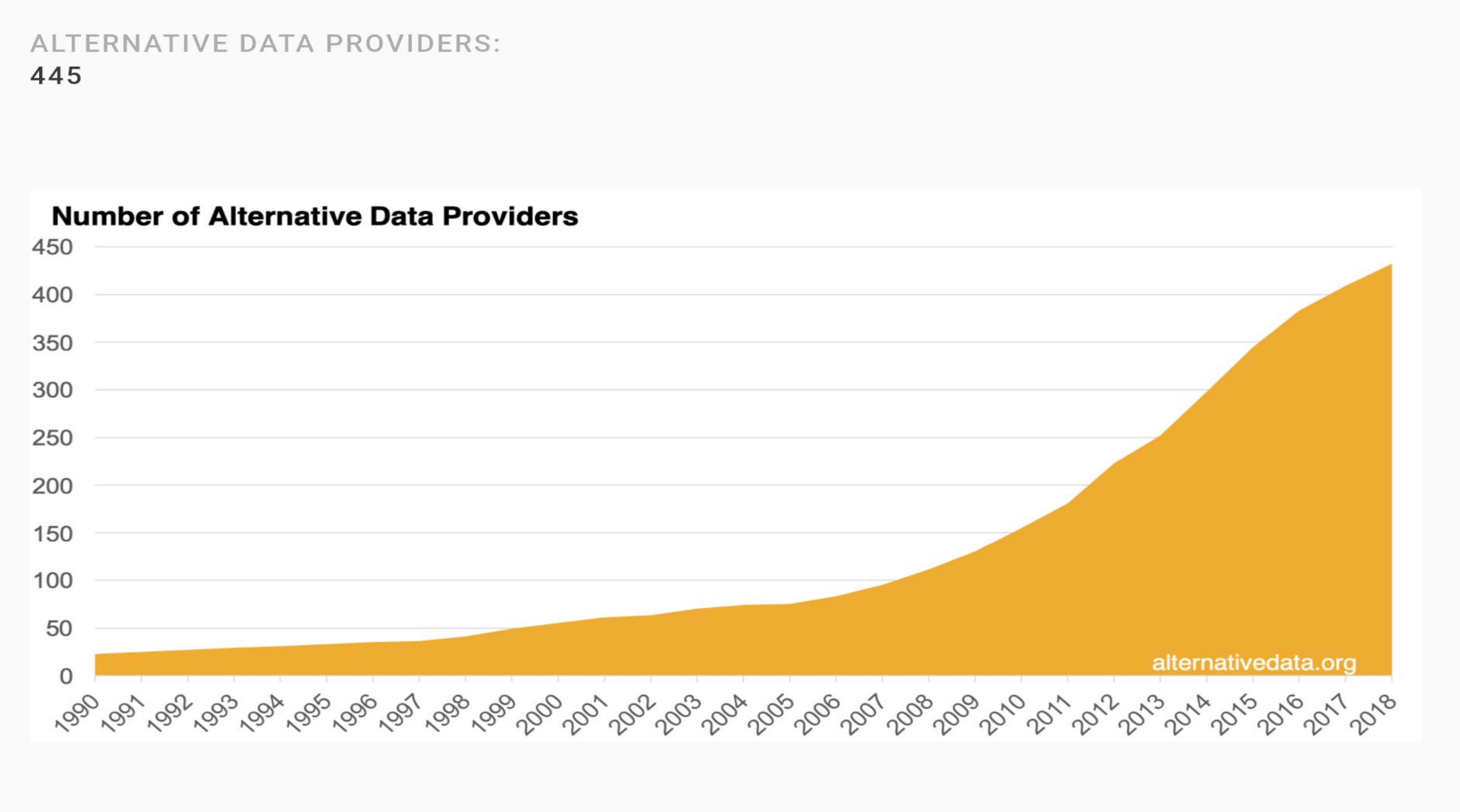


DOMO
DATA NEVER SLEEPS 8.0
How much data is generated every minute?
2020
every
MINUTE
of the
DAY
ZOOM
HOSTS 208,333
PARTICIPANTS IN MEETINGS
REDDIT
SEES
479,452
PEOPLE ENGAGE WITH CONTENT
NETFLIX
USERS STREAM
404,444
HOURS OF VIDEO
DOORDASH
DINERS ORDER
555
MEALS
INSTAGRAM
USERS POST
347,222
STORIES
YOUTUBE
USERS UPLOAD
500 HRS
OF VIDEO
TWITTER
GAINS NEW USERS 319
CONSUMERS SPEND
$1,000,000
ONLINE
MICROSOFT TEAMS
CONNECT
52,083
USERS
FACEBOOK
USERS SHARE
150,000
MESSAGES
LINKEDIN
USERS APPLY FOR
69,444 JOBS
TIKTOK
IS INSTALLED
2,704 TIMES
VENMO
USERS SEND
$239,196
WORTH OF PAYMENTS
SPOTIFY
ADDS
28 TRACKS
TO ITS MUSIC LIBRARY
INSTAGRAM
BUSINESS PROFILE ADS
SEE CLICKS 138,889
AMAZON
6,659 SHIPS
PACKAGES
1,388,889
PEOPLE MAKE VIDEO/VOICE CALLS
$3,805 IS SPENT ON MOBILE APPS
WHATSAPP
USERS SHARE
41,666,667
MESSAGES
FACEBOOK
USERS UPLOAD
147,000
PHOTOS
3.0
3.4
4.3
4.5
2014
2016
2018
2020
GLOBAL INTERNET POPULATION GROWTH 2014–2020
Learn more at domo.com

Alternative data

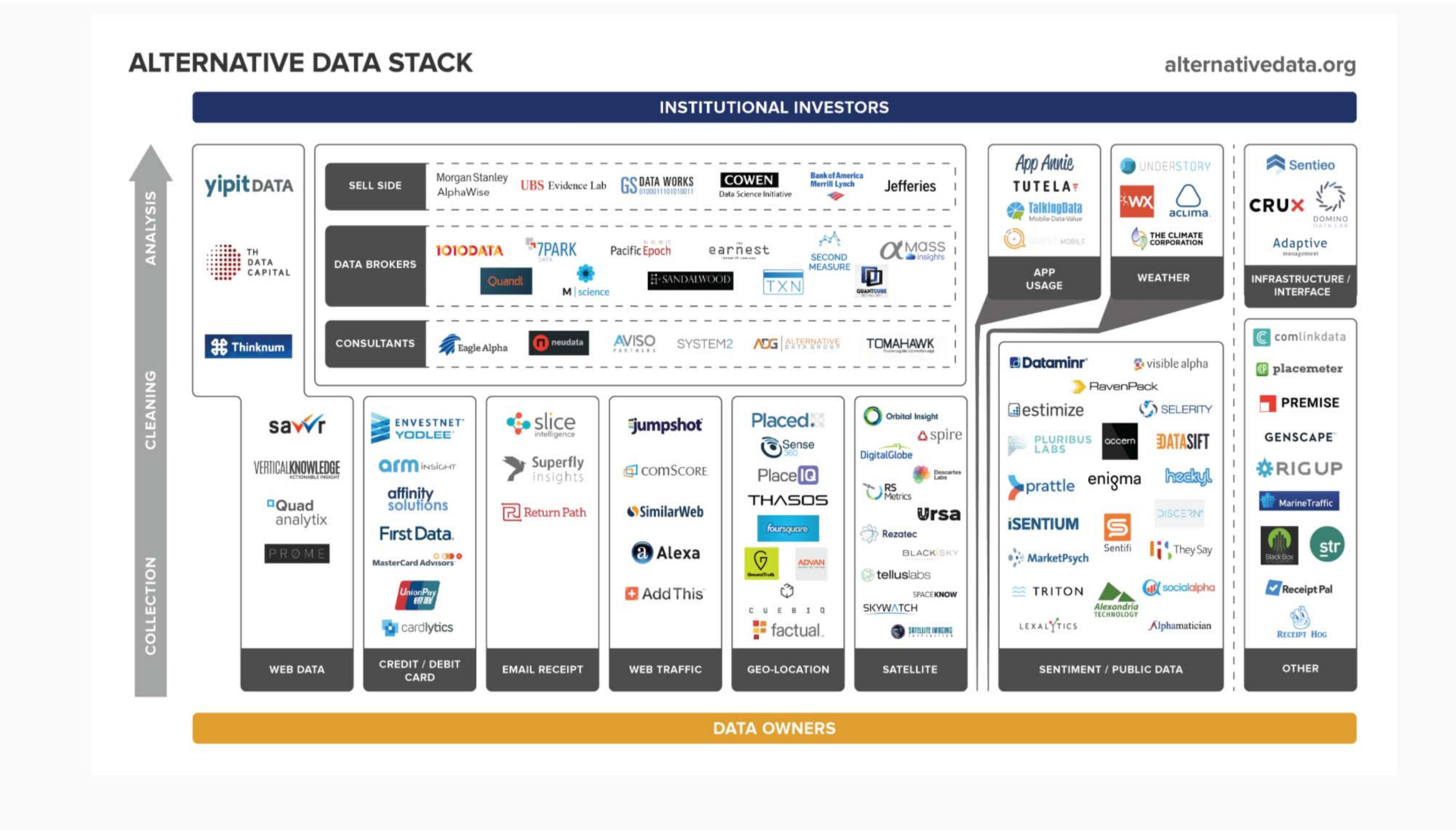
- **Examples:**
  - Social media (Tweeter, Facebook, StockTwits, Estimize, etc; what people think)
  - Geolocation data (where people shop)
  - Credit card and Point of Sales (POS) data (what people buy)
  - Satellite Imagery (e.g., parking lots fill rates at retailers)
  - Search traffic on the internet (e.g., clickstream data, google searches; what people are paying attention to)

New data vendors



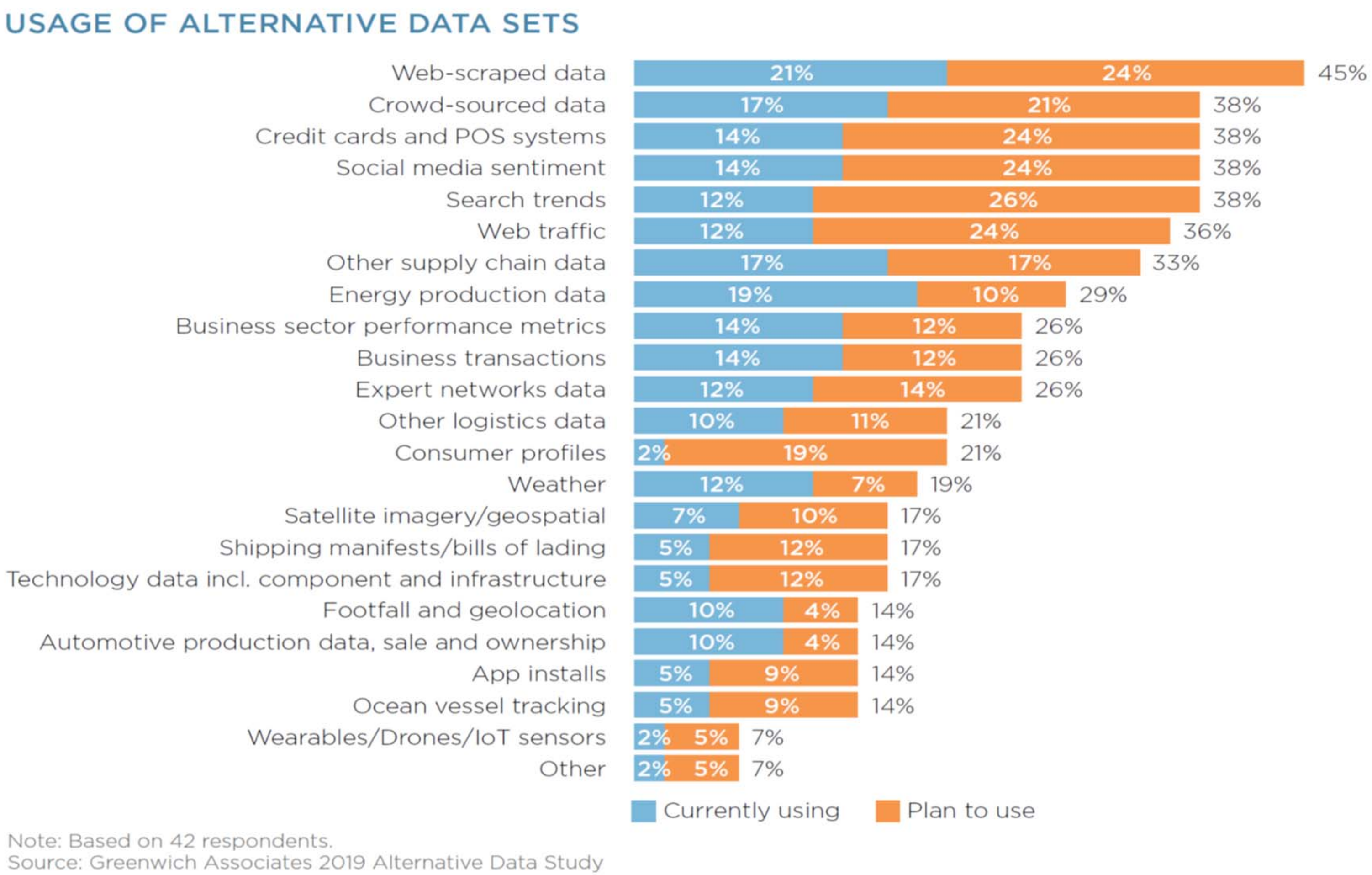
Source: AlternativeData.org

The new information intermediaries



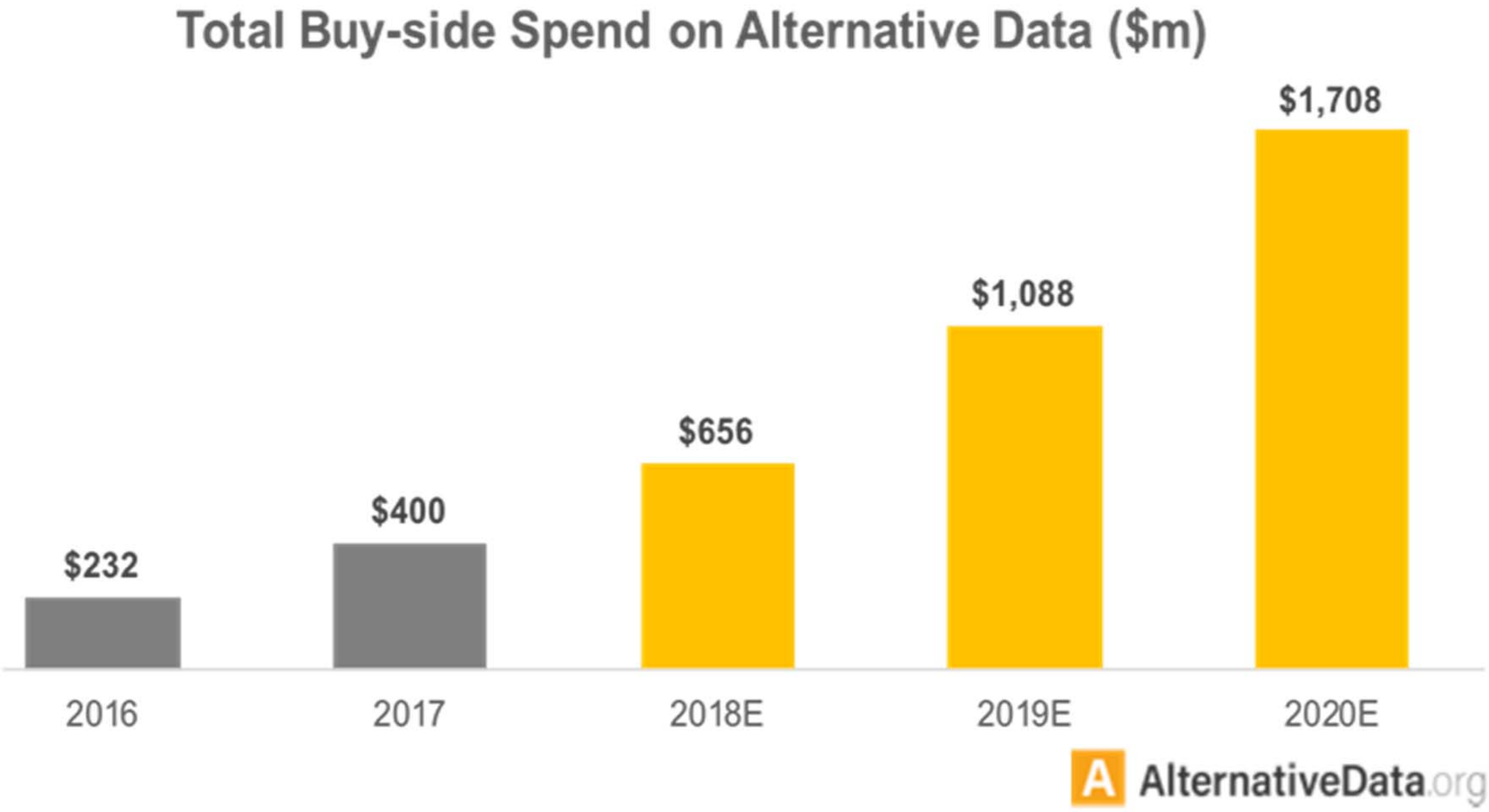
Source: AlternativeData.org

Alternative data



Source: "Demystifying Alternative Data"-Greenwich Associates, 2019

Investment in alternative data



Source: AlternativeData.org

Questions

- This evolution raises many interesting (and challenging) questions:
  - Does it improve financial forecasting? At which horizon (ST/LT/Both)?
  - Should one rely only on humans, machines, a combination of both for financial forecasting?
  - Is there information about fundamentals/returns in alternative data? At which horizon?
  - How does this evolution affect active asset managers (rise of the "quants")
  - Do alternative data make securities markets more or less informative about fundamentals (prices closer to fundamentals)?
- Research/knowledge on these questions is limited

Research Challenges

- **Challenge 1:** Investors' forecasts are difficult to observe. One way to address this problem is to consider (sell-side) analysts' earnings forecasts:
  - There is evidence that stock analysts rely on alternative data to formulate their forecasts
  - They must formulate forecasts at various horizons to formulate "price targets"/Buy/Sell Recommendations
  - The quality of their forecasts matter for their careers
  - Their forecasts/recommendation move stock prices
- **Challenge 2:** There are many different types of Alternative Data (may not be equally informative or informative about the same horizon).

Analysts and Alternative Data

["Nowadays, analysts sift through non-traditional information such as satellite imagery and credit card data or use machine learning and natural language processing to glean fresh insights from traditional sources such as economic data and earnings call transcripts”](in “How investment analysts became data miners” in Financial Times, November 28, 2019)"

Evidence from analysts' reports

Wal-Mart Stores 12 August 2010

UBS Proprietary National Parking Lot Fill Rate Analysis

We have conducted an analysis with Remote Sensing Metrics, LLC to track parking lot fill rates in order to predict overall US comp-sales performance at Walmart Stores using a sample of between 100 and 150 like-for-like satellite images each month for the past six months. Samples are representative of geographic region, store formats, day of week, and the time of period analysis. All satellite images are usually taken between 10:30am and 1pm to minimize shadows on the images. We believe a traditional grocery trip is less fixed to a certain time of day and thus the time-slot window for imagery results bears less risk than for other more discretionary shopping trips.

Table 2

Number of Analyst Forecasts Explicitly Supported by Data from a given Alternative Data Category

In this table we present the numbers of analyst forecasts that are explicitly supported by alternative data by alternative data category. Since a given analyst report may draw from multiple alternative data categories, the sum of the number of forecasts in Table 2 exceeds the total number of forecasts explicitly supported by alternative data reported in Table 1.

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

Chi, Wang and Zeng (2021), "The use and usefulness of big data in finance: Evidence from Financial Analysts", Working Paper, Cornell University

Does Alternative Data Improve Financial Forecasting? The Horizon Effect

Olivier Dessaint (INSEAD), Thierry Foucault (HEC Paris), and Laurent Frésard (Université de Lugano) (2021)

(available here)

Overview

- We use a large sample of sell-side analysts' earnings forecasts at various horizons (up to 5 years) obtained from I/B/E/S from 1983 to 2017

- We study (i) the long run evolution of the informativeness of these forecasts and (ii) how it changes when new social media data (StockTwits) becomes available.

Measuring analysts' forecasts informativeness

- On each day t and for each analyst i in our sample, we:
  - Retrieve latest forecast $\boldsymbol{e^{f}_{jh}}$ and realization $\boldsymbol{e^{a}_{jh}}$ for each covered stock j at horizon h (median #days until $\boldsymbol{e^{a}_{jh}}$ is publicly announced)
  - Regress realized earnings on earnings forecasts (normalized by "Asset Size") for all stocks j in the analysts' portfolio

$$\frac{e^{a}_{j\mathrm{h}}}{Assets_j} = \alpha + \beta \frac{e^{f}_{j\mathrm{h}}}{Assets_j} + \varepsilon_j$$

  - Estimate $\mathbf{R^2_{i,t,h}}$ of the regression (proportion of the variance of realized earnings explained by the analyst's forecast);
  - Similar to estimating the analyst's **average forecasting error** at a given horizon but controlling for the dispersion of the forecasted variable.

Measuring Forecast Informativeness across horizons

- Example: Analyst Esther covers 6 stocks at time t = 12/31/2006

Step 1 Identify forecasted fiscal period

Step 2 Identify last available earnings forecasts for A, B, C, .., F

Step 3 Retrieve actual earnings

Firm covered by Esther	Latest Estimates (f)	Realised Earnings (θ)
	Fiscal Period = 12/31/2008	
A	$110M	$120M
B	$35M	$25M
C	$65M	$55M
D	$210M	$150M
E	$105M	$90M
F	$150M	$100M

Measuring Forecast Informativeness across horizons

- Example: Analyst Esther covers 6 stocks at time t = 12/31/2006

Firm covered by Esther	Latest Estimates (f)	Realised Earnings (θ)
	Fiscal Period = 12/31/2008	
A	0.10	0.06
B	0.12	0.07
C	0.08	0.02
D	0.11	0.08
E	0.10	0.12
F	0.12	0.06

Step 4
Normalize
by total assets

Step 4 Normalize by total assets

Measuring Forecast Informativeness across horizons

- Example: Analyst Esther covers 6 stocks at time t = 12/31/2006

Firm covered by Esther	Latest Estimates (f)	Realised Earnings (θ)
	Fiscal Period = 12/31/2008	
A	0.10	0.06
B	0.12	0.07
C	0.08	0.02
D	0.11	0.08
E	0.10	0.12
F	0.12	0.06

Step 5
Regress θ on f and estimate R_t^2

- Informativeness: $R_t^2 = 15\%$

Measuring Forecast Informativeness across horizons

- Example: Analyst Esther covers 6 stocks at time t = **12/31/2006**

Step 6
Retrieve actual earnings announcement date

Firm covered by Esther	Latest Estimates (f)	Realised Earnings (θ)	Earnings (θ) report date
	Fiscal Period = 12/31/2008		
A	0.10	0.06	31/03/2009
B	0.12	0.07	31/03/2009
C	0.08	0.02	31/03/2009
D	0.11	0.08	31/03/2009
E	0.10	0.12	31/03/2009
F	0.12	0.06	31/03/2009

Step 7
Compute horizon as # mths between t and (median) report date

- Informativeness: $R_t^2 = 15\%$
- Horizon: # months between 12/31/2006 and 31/03/2009 = 27 months

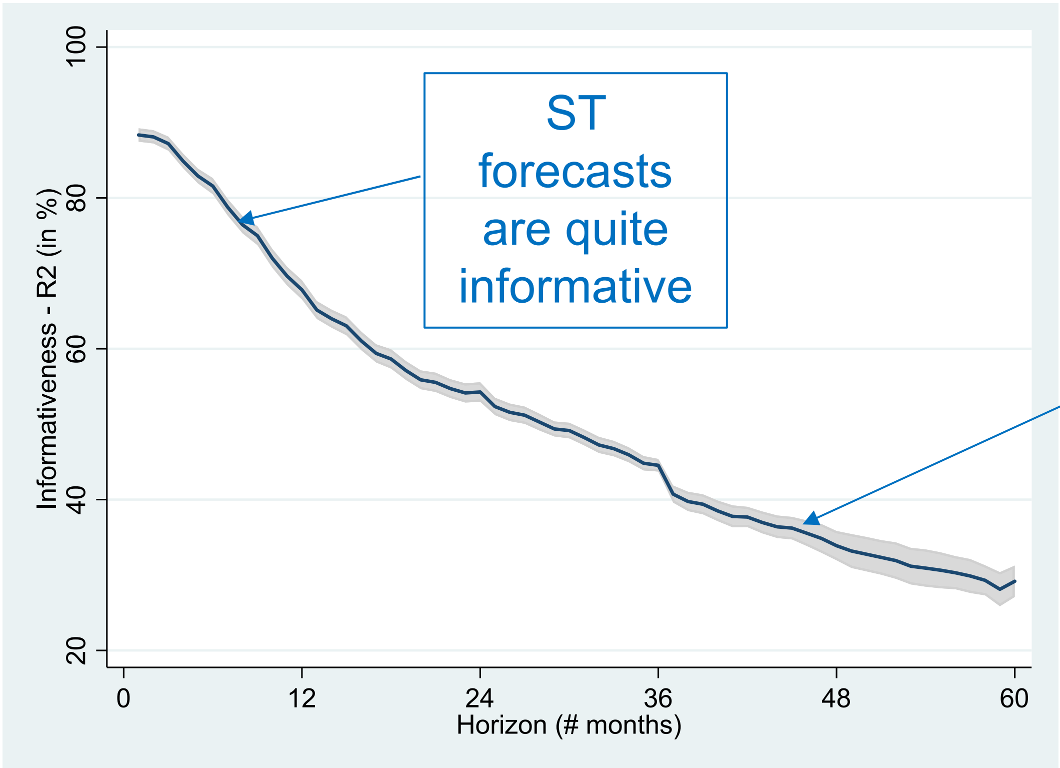
Long Run Evolution

Term Structure of Analysts' Forecasts Informativeness

Sample of 65, mio obs. of R2 by analyst-day-horizon

From 1983 to 2017

Average R2 decreases with horizon



LT forecasts are much less informative

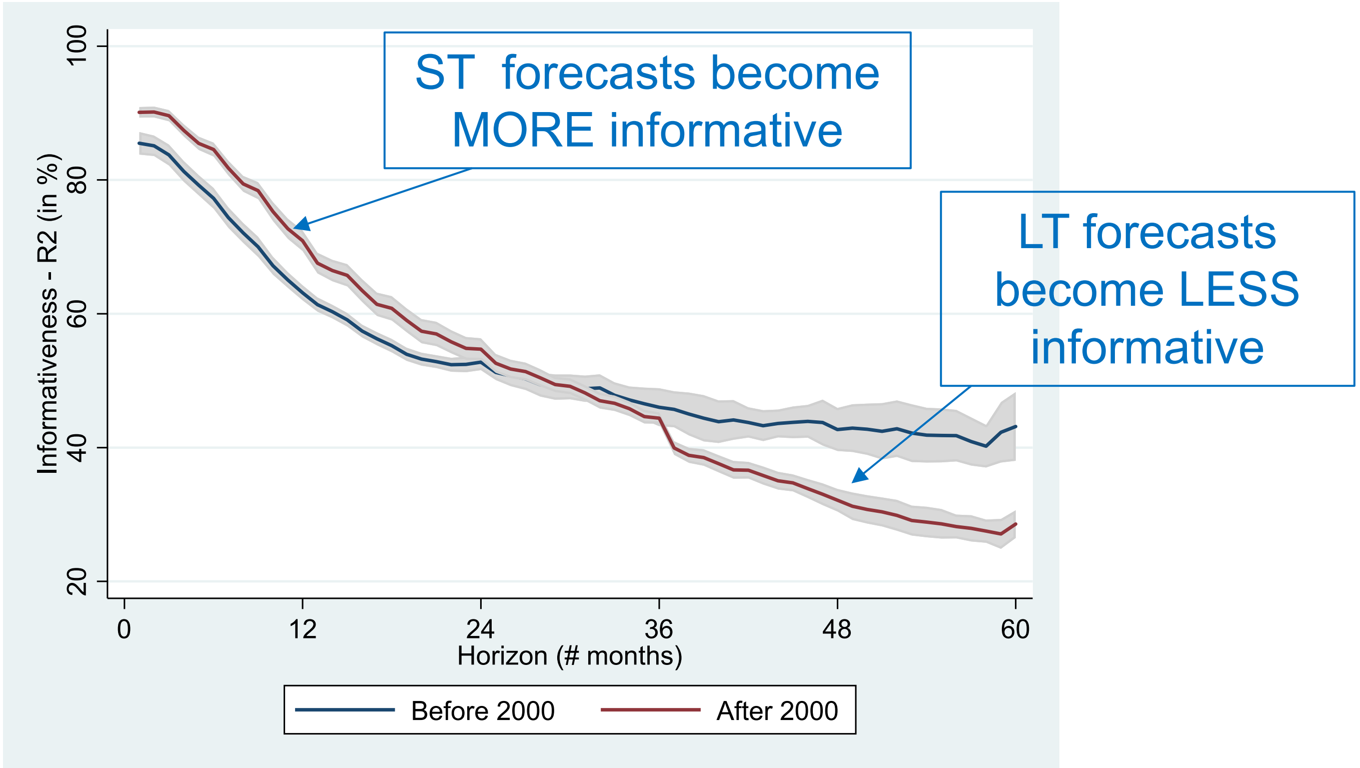
Summary Stats

	N	Mean	STDV	Min	P25	P50	P75	Max
Sample: Horizon <=1 Yr								
Forecast informativeness - R2 measure	33'413'667	79.60	27.63	0.00	72.57	92.49	98.42	100.00
Horizon	33'413'667	0.49	0.29	0.00	0.24	0.49	0.74	1.00
#Stocks Covered	33'413'667	8.29	5.36	3.00	4.00	7.00	11.00	30.00
Sample: 1 Yr <= Horizon <2 Yrs								
Forecast informativeness - R2 measure	25'060'925	59.21	34.64	0.00	29.37	69.51	90.42	100.00
Horizon	25'060'925	1.45	0.28	1.00	1.21	1.43	1.68	2.00
#Stocks Covered	25'060'925	8.14	5.09	3.00	4.00	7.00	11.00	30.00
Sample: 2 Yrs <= Horizon <3 Yrs								
Forecast informativeness - R2 measure	5'361'069	49.37	36.23	0.00	10.47	53.15	84.34	100.00
Horizon	5'361'069	2.39	0.28	2.00	2.15	2.34	2.61	3.00
#Stocks Covered	5'361'069	7.53	4.71	3.00	4.00	6.00	10.00	30.00
Sample: 3 Yrs <= Horizon <4 Yrs								
Forecast informativeness - R2 measure	1'349'749	37.62	36.04	0.00	0.00	28.84	71.60	100.00
Horizon	1'349'749	3.45	0.29	3.00	3.20	3.43	3.70	4.00
#Stocks Covered	1'349'749	6.70	3.95	3.00	4.00	6.00	9.00	30.00
Sample: 4 Yrs <= Horizon <5 Yrs								
Forecast informativeness - R2 measure	703'050	31.18	34.98	0.00	0.00	14.75	62.31	100.00
Horizon	703'050	4.43	0.28	4.00	4.19	4.39	4.65	5.00
#Stocks Covered	703'050	6.26	3.54	3.00	4.00	5.00	8.00	30.00

Evolution Over Time

The term-structure becomes steeper post-2000

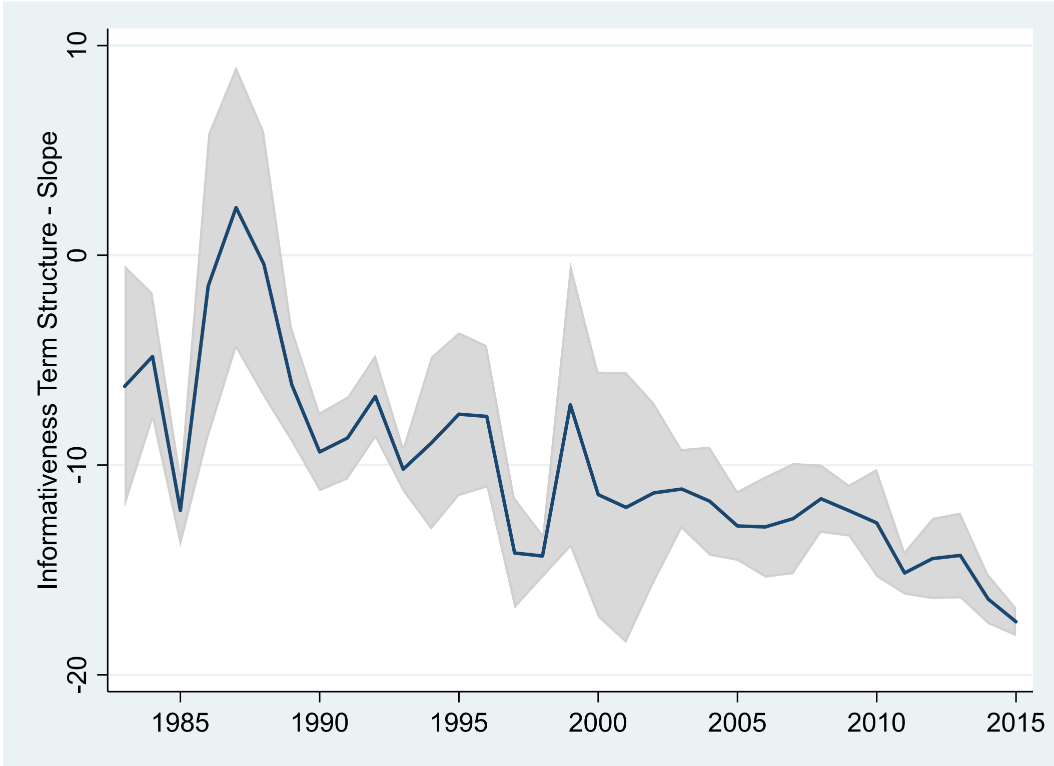
Does not depend on where we cut (2000 is not important in itself)



Evolution over time

The slope of the term structure of analysts' forecast informativeness becomes steeper over time

With an **acceleration after 2005**



Estimate the slope of the term stucture **every year t** by regressing $R^2_{i,t,h}$ on horizon *h*

$$R^2_{i,t,h} = \alpha + \beta \times h + \varepsilon$$

Forecast informativeness over time

- More formal: regress $R^2_{i,t,h}$ on a annual time trend by horizon

Panel A - All analysts

	Dependent variable: Forecast informativeness (R2 measure in percentage points)									
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample	h <1 Yr		1 Yr <= h <2 Yrs		2 Yrs <= h <3 Yrs		3 Yrs <= h <4 Yrs		4 Yrs <= h <5 Yrs	
Normalized Year Trend	11.5***	11.0***	9.4***	8.4***	2.4	0.3	-11.5***	-7.2***	-20.0***	-13.9***
	(8.00)	(7.78)	(6.89)	(6.07)	(1.46)	(0.20)	(-5.12)	(-2.75)	(-5.42)	(-3.39)
Constant (1983-1992)	74.7***		55.0***		47.9***		44.3***		42.6***	
	(93.81)		(82.46)		(39.10)		(29.78)		(21.12)	
SIC2 FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Size Quintile FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Age FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	33'413'667	31'308'798	25'060'925	23'326'180	5'361'069	5'012'427	1'349'749	1'291'499	703'050	672'490

cluster S.E. by forecasting period

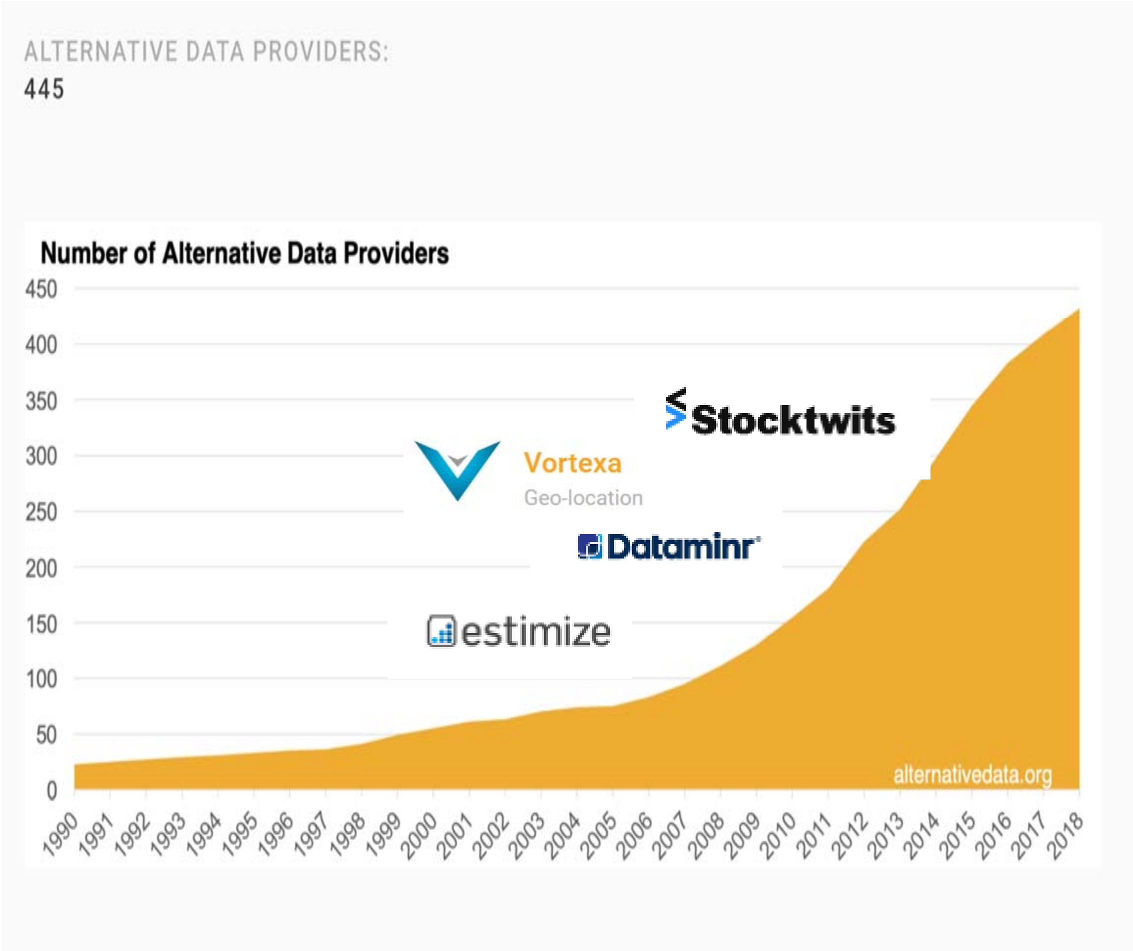
ST forecasts become MORE informative

LT forecasts become LESS informative

Explanations?

Many possible factors can explain this trend in the term structure of analysts' **forecasts informativeness.**

We conjecture that one factor is the growth in alternative data and present evidence and theory consistent with this conjecture.



Theory (sketch; more details in the paper)

- Forecasting LT and ST cash-flows about a firm are related but distinct tasks:
  - Related: Fixed cost in understanding a particular firm/industry.
  - Different: The nature of information useful to forecast LT cash flows is different from the nature of information useful to forecast ST cash-flows.
  - For various reasons (cost, agency etc.), splitting these tasks between different individuals is not easy
- Alternative data is mainly informative about short-term fundamentals, not long term.
  - Consistent with empirical evidence on information in social media for instance

Theory (sketch; more details in the paper)

- Forecasters (e.g., analysts) must therefore choose how much effort to allocate to each task.
  - Multitasking is costly: More effort allocated to one task makes effort allocated to the other task more costly.
- If the cost of effort (or the return on effort) for obtaining and processing ST information declines, forecasters optimally allocate more effort to this task and less to the other task (intuition: agents equalize marginal benefits of effort on each task).
  - Consequence: The quality/informativeness of ST forecasts improves but that of LT forecasts declines

Test

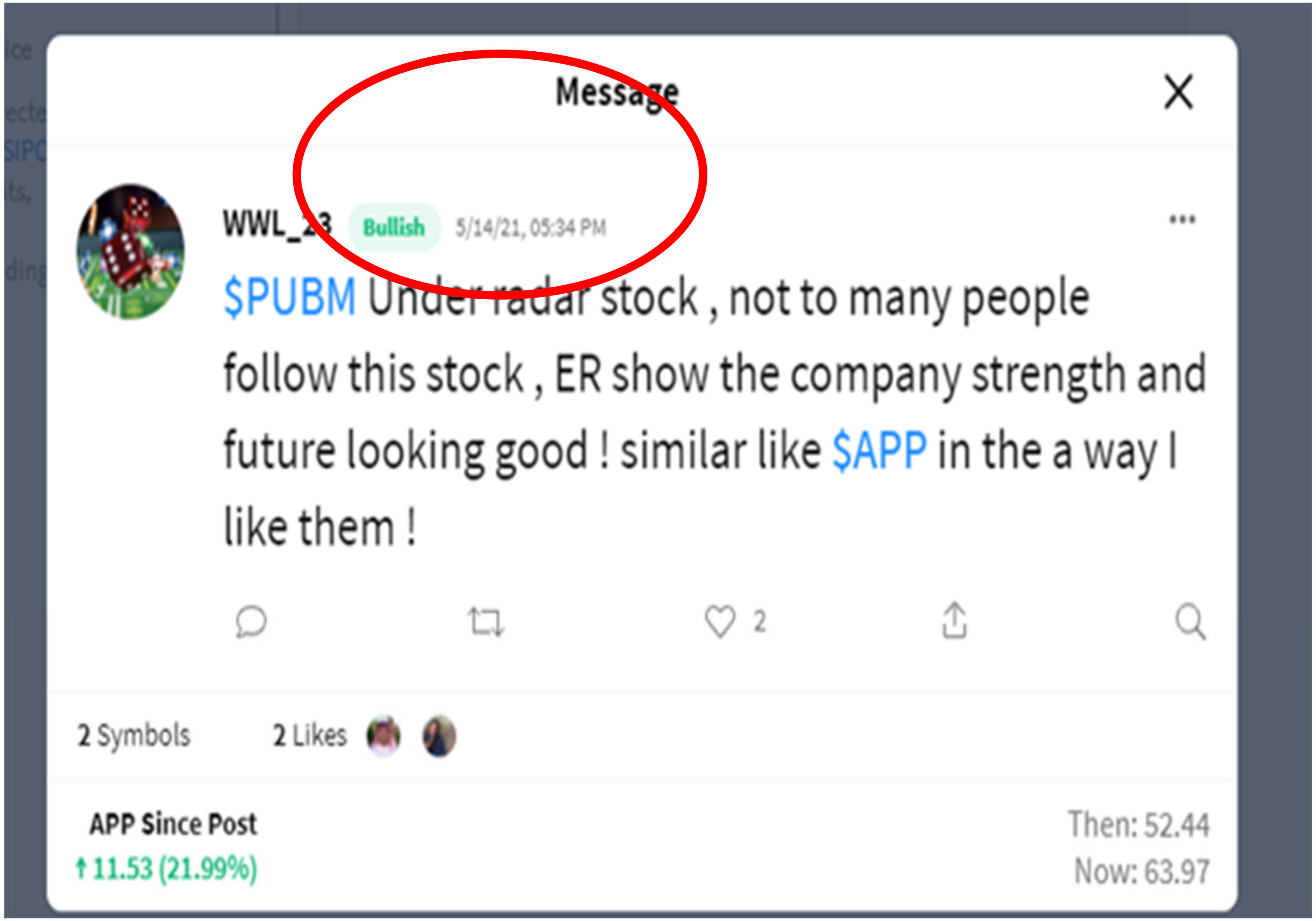
- To test this prediction, we consider a shock on the alternative data available to analysts
  - The introduction of a new social media ("StockTwits") dedicated to equity markets in the U.S.
  - This shock expands the set of data that can be used by analysts to form their forecasts (and we show that they do rely to some extent on these data).
- We test whether:
  - The term structure of forecasts informativeness changes as predicted for analysts who are the most exposed to the shock.

StockTwits and Analysts' Forecasts Informativeness

StockTwits

- StockTwits is the largest social network fully dedicated to US financial markets
  - Founded in 2008
  - Discussion platform like "Twitter" for investors
  - Traders create posts with charts, links to articles, and opinions about stocks
  - Posts are linked to securities via a “cashtag”: $+TICKER (ex. $F refers to Ford)
  - 2 million users by mid-2019 / More than 4 million messages per month

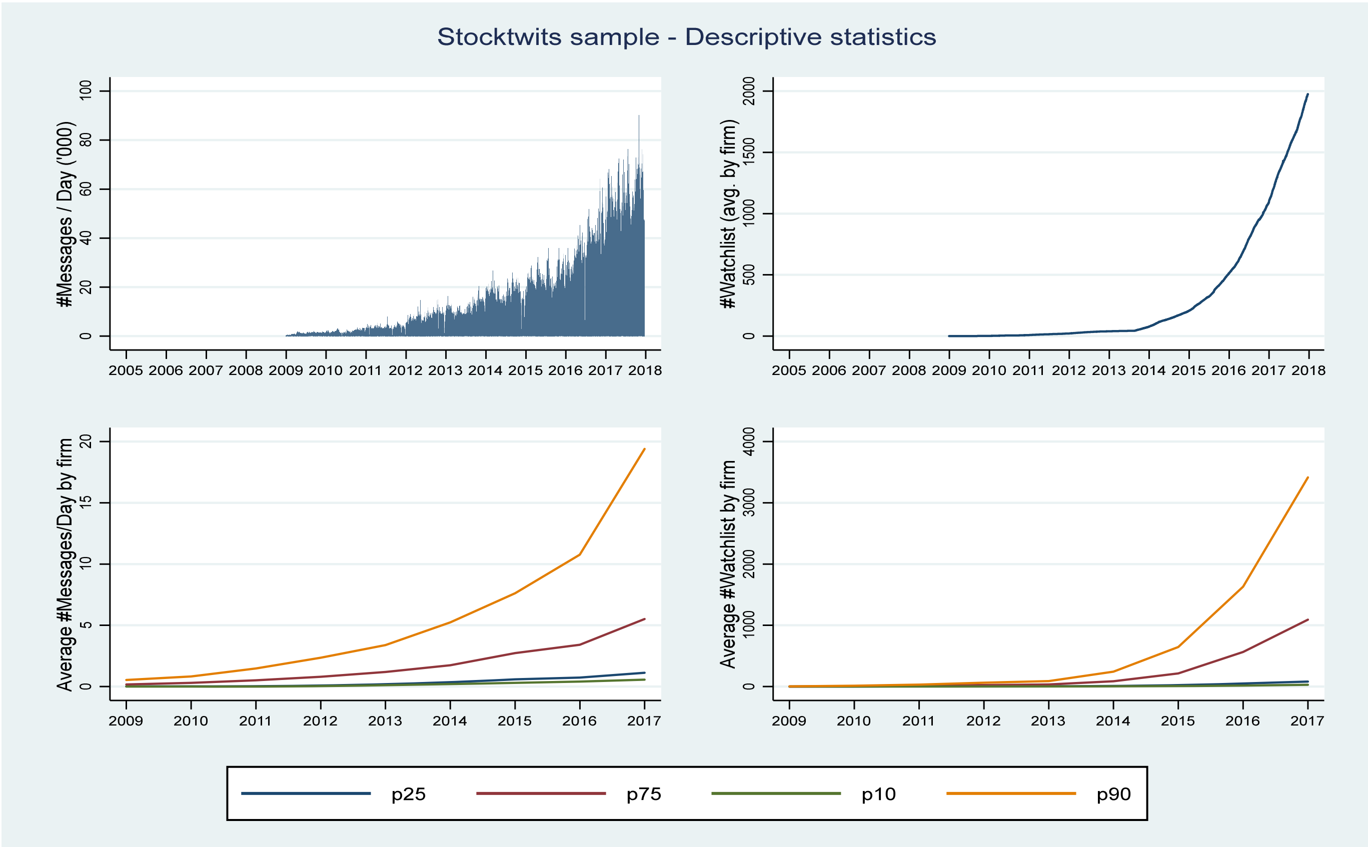
StockTwits-Example



StockTwits Sample

- StockTwits granted us the access to all messages posted on the platform from 2009 to 2017.
  - This dataset includes more than 20 million messages about U.S. listed stocks and contains information about
    - The message: date created, source, text
    - The underlying asset: ticker, name, #watchlist, industry, trending, …
    - The user who wrote the message: name, background, experience, location, number of followers, self-declared trading style,…
- Analysis starts in 2005, i.e. 5 years before StockTwits plateform takes-off
  - All indicators of social media activity from StockTwits are set to zero before 2009

StockTwits Expansion



Empirical Design

- Our test: Use the introduction and expansion of StockTwits as a source of variation in alternative data available to analysts
  - Idea: capture variation in data only generated by the existence of the platform (that would not be available otherwise, e.g., news)
- "Treatment": A given analyst's exposure to StockTwits
  - Amount of data about firms covered by that analyst
  - Exposure varies over time (expansion) and across analysts (distinct coverage)
- Main prediction:
  - More exposure to StockTwits increases short-term informativeness (R^2)
  - More exposure to StockTwits decreases long-term informativeness (R^2)

Measure of Analysts' Exposure to StockTwits

- Two measures for a given analyst exposure on day *t*:

1. ***#Watchlist* aggregated across firms covered**
   - StockTwits users report a private list of all the firms they watch
   - *#Watchlist* is the number of users that have a given firm on their watchlist
   - Users rarely modify their watchlist --> **variation comes from expansion**
   - **Orthogonal to news arrival (from other sources)** ✓

2. **<u>Hypothetical</u> messages (30 days) aggregated across firms covered**
   - \# *Actual Messages* correlates with news arrival ✗
   - Hypothetical messages for firm i= Total #messages on StockTwits x historical "market share" of all messages by firm i.
   - The message market share is persistent --> **variation in exposure comes from expansion**
   - **Orthogonal to news arrival (from other sources)** ✓

Is StockTwits a good laboratory?

- To capture variation in cost of producing ST info from data, our measures of exposure based on StockTwits needs to:

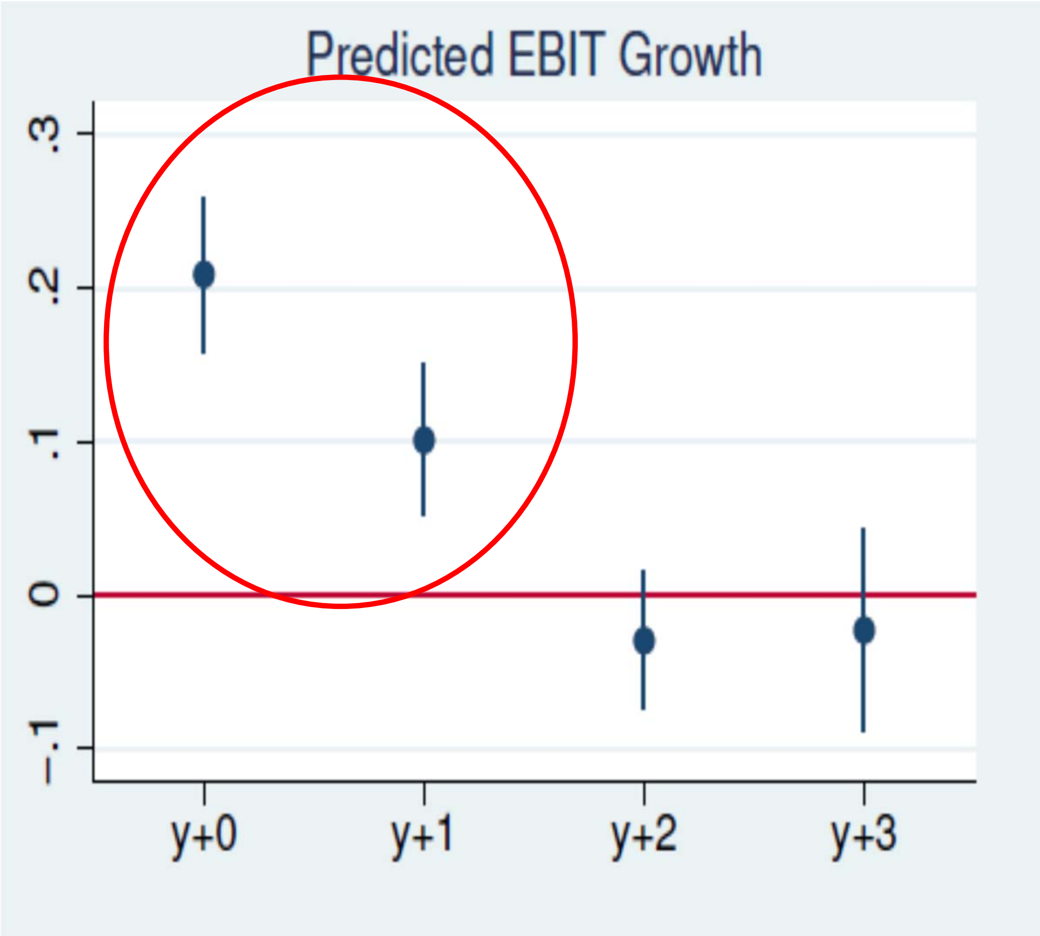
  1. Contain information mainly relevant about **short-term** earnings

  2. Used by financial analysts

We believe both conditions are likely to hold

- Existing research indicates that social media data contains info relevant for predicting short-term outcomes (e.g., sales)
  - Chen, De, Hu, and Hwang (2014) or James, Johnston, Markov and Wolfe (2016)

There is information in StockTwits



Cross-sectional regressions of year-on-year EBITGrowth (up to 3 years) on "Ratings" (the difference between the fraction of Bearish and Bullish messages (as declared by users) about a firm in StockTwits.

Regressions are done per quintile. The graphic reports the average coefficient on 'Ratings" per quintile.

Do analysts follow StockTwits?

- StockTwits has been integrated into all major aggregation plateforms likely used by analysts
  - Bloomberg.com, Reuters.com, CNN Money, or Yahoo!
  - Some use robo-analysts to help analysts filter information

- Analysts are more likely to make (or revise) forecasts on a firm and day following more intense activity on StockTwits
  - This result holds when we control for trading volume
  - Even when no news is announced in the past 30 days

- Use biographic information about analysts to match analysts with StockTwits' user accounts (based on names)
  - 35% of exact matches (based on 7,656 analysts during 2009-2017 period)

Main specification

- The **treatment**: Exposure to social media data generated by StockTwits
  - All tests are at the analyst level
- We estimate for different horizons h:

$$R^2_{i,t,h} = \alpha_{i,h} + \alpha_{t,h} + \boldsymbol{\lambda_h}\ \boldsymbol{Data\ Exposure_{i,t-1}} + Controls + \varepsilon_{t,j}$$

  - Analyst and day fixed effects (within-analyst variation over time)
  - Data Exposure=0 before 2009 (creation of the platform)
  - **λ_h measures the effect of exposure to social media data on the average analyst forecast informativeness for horizon h**

Findings

- Exposure to social media data affects differently the informativeness of analysts short term and long term forecasts

Panel A - Proxy for Social Media Data : # Watchlist

	Dependent variable: Forecast informativeness (R2 measure in percentage points)							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	h <=1 Yr		1 Yr < h <=2 Yrs		2 Yrs < h <=3 Yrs		h>=3 Yrs	
Data Exposure	0.54***	0.53***	0,4	0,18	-0.65***	-1.00***	-1.51***	-1.55***
	(3.90)	(4.03)	(1.07)	(0.47)	(-3.20)	(-4.78)	(-3.49)	(-3.20)
Analysts FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	14.026.800	13.006.543	11.502.199	10.612.608	3.929.446	3.648.151	1.500.165	1.438.756

cluster S.E. by forecasting period

ST forecasts become MORE informative

LT forecasts become LESS informative

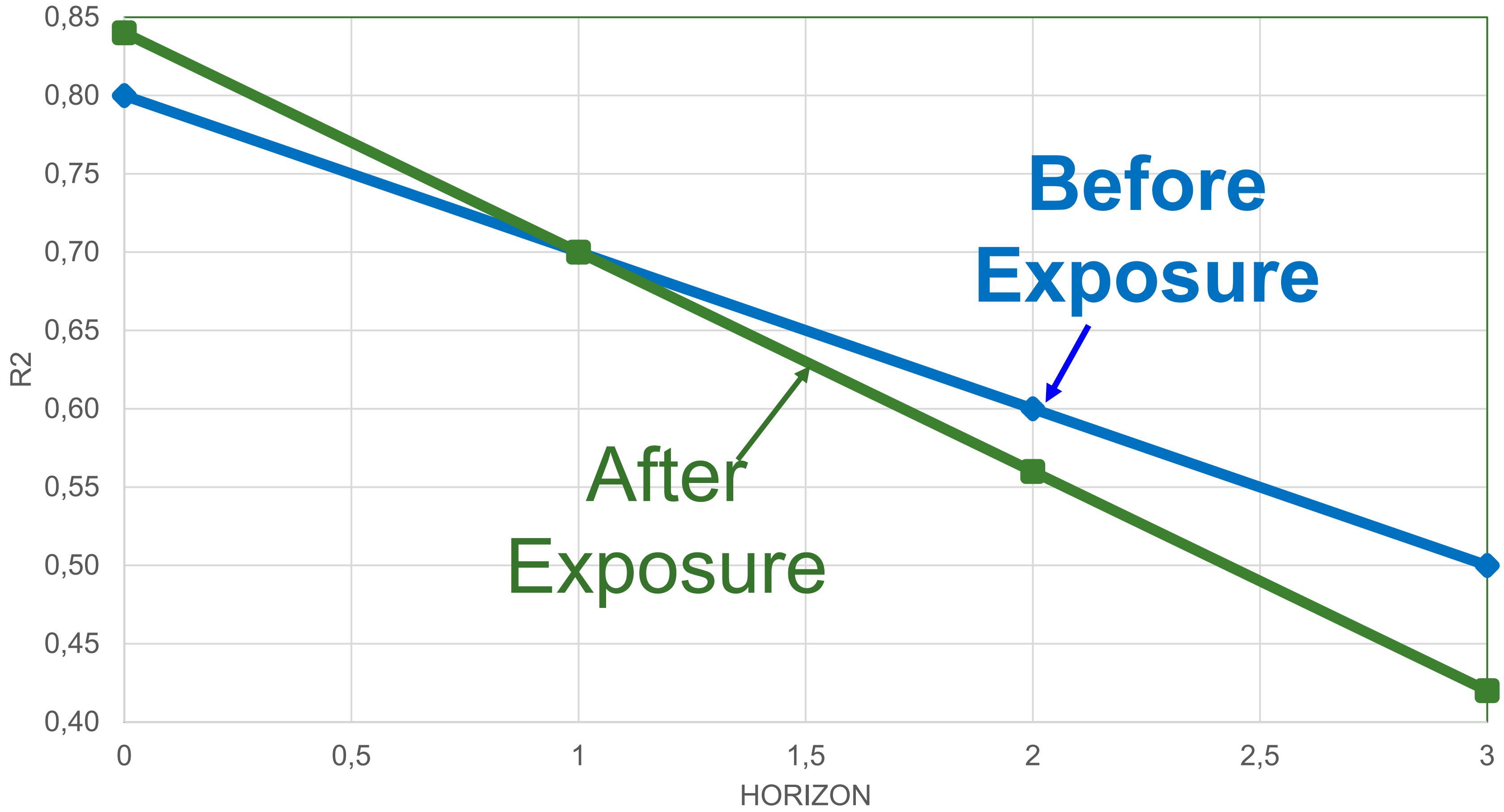
Findings

- Alternative: $R^2_{i,t,h} = \alpha_i + \alpha_t + \lambda_0(h-1) + \lambda_1 Exposure_{i,t} + \lambda_2\,(h-1) \times Exposure_{i,t} + \varepsilon_{i,t}$

Dep. variable:	Forecast informativeness (R^2)					
Data Exposure Proxy:		*#Watchlist*			*#Hypothetical Messages*	
OLS:	(1)	(2)	(3)	(4)	(5)	(6)
$h^* \times$ Data Exposure	-0.86***	-0.78***	-0.96***	-0.69***	-0.94***	-1.05***
	(-2.59)	(-3.06)	(-3.72)	(-2.75)	(-4.54)	(-5.03)
Data Exposure	0.13	-0.17	-0.35	0.34	-0.14	-0.32
	(0.50)	(-0.64)	(-1.29)	(1.42)	(-0.57)	(-1.30)
h^*	-16.66***			-16.62***		
	(-33.85)			(-32.13)		
Analyst FE	Yes			Yes		
Date FE	Yes			Yes		
Analyst FE (interacted)		Yes	Yes		Yes	Yes
Date FE (interacted)		Yes	Yes		Yes	Yes
Controls			Yes			Yes
N	30,959,281	30,105,556	27,860,429	30,959,281	30,105,556	27,860,429

$\lambda_2 < 0$: Exposure to social media data renders LT forecast less informative

Average effect of analysts' exposure to StockTwits



Exposure= Stocks covered by an analyst starts being covered on StockTwits

Additional Results

- Deterioration in the informativeness of LT forecasts for an analyst **<u>should be</u>** more pronounced when (followed from theory):

  - The cost of multitasking is larger for an analyst

    - We use number of stocks followed by an analyst as a proxy for this cost

  - Earnings are less auto-correlated (ST info is less useful to predict LT earnings)

Multitasking Costs

- We use number of stocks followed by analysts as proxy for task multiplicity

Dependent variable: Forecast informativeness (R2 measure in percentage points)

OLS	(1)	(2)	(3)	(4)	(5)	(6)
Proxy for Social Media Data		# Watchlist			# Hypothetical Messages Last 30 days	
Sample		Full			Full	
Horizon x Data Exposure x #Firms	-0.14***	-0.06***	-0.06***	-0.10***	-0.05*	-0.06***
	(-5.71)	(-3.38)	(-3.82)	(-6.81)	(-1.94)	(-2.65)
Analysts FE	Yes			Yes		
Date FE	Yes			Yes		
Analysts x Horizon FE		Yes	Yes		Yes	Yes
Date x Horizon FE		Yes	Yes		Yes	Yes
Controls			Yes			Yes
N	30.958.705	30.105.299	27.860.178	30.958.705	30.105.299	27.860.178

cluster S.E. by forecasting period

  - Triple interaction: deterioration of LT informativeness greater for analysts covering more stocks

Correlated earnings

- We use quarterly earnings AR(1) coefficient (two-years rolling windows) aggregated across firms covered by analysts

Dependent variable: Forecast informativeness (R2 measure in percentage points)

OLS	(1)	(2)	(3)	(4)	(5)	(6)
Proxy for Social Media Data	# Watchlist			# Hypothetical Messages Last 30 days		
Sample	Full			Full		
Horizon x Data Exposure x Auto-correlation	1.17***	0.64***	0.58***	0.69***	0.39**	0.35**
	(3.23)	(2.82)	(2.62)	(2.92)	(2.44)	(2.35)
Analysts FE	Yes			Yes		
Date FE	Yes			Yes		
Analysts x Horizon FE		Yes	Yes		Yes	Yes
Date x Horizon FE		Yes	Yes		Yes	Yes
Controls			Yes			Yes
N	28.711.790	27.865.669	27.840.732	28.711.790	27.865.669	27.840.732

cluster S.E. by forecasting period

  - Triple interaction: deterioration of LT informativeness smaller for analysts covering stocks with more auto-correlated earnings

Conclusion

- **Does alternative data improve financial forecasting?**
  - Yes at short horizons
  - No at long horizons
  - Even worse it seems to impair the quality of analysts' forecasts at long horizons
- **Does alternative data make financial markets more informationally efficient/more informative?**
  - There is academic evidence that this is the case at relatively short horizons (< one year)
  - Maybe not at long horizons…We do not know. Difficult question. Measuring the informativeness of stock prices at various horizons is tough….

Next Steps

- Implications for the informativeness of stock prices?
  - In theory, there is a positive relationship between the informativeness of asset prices about fundamentals and a weighted average of investors' forecasts informativeness (Foucault, Dessaint and Frésard (2021), work in progress).
  - Have asset prices become more short-termist because forecasters are more short-termists?

- Implications for corporate investment?
  - Are firms with long horizons investment projects penalized? (Larger cost of capital/Lower investment)