

On EDGAR Log Database

Present by: Long Zhen

2023.10.24

What is EDGAR

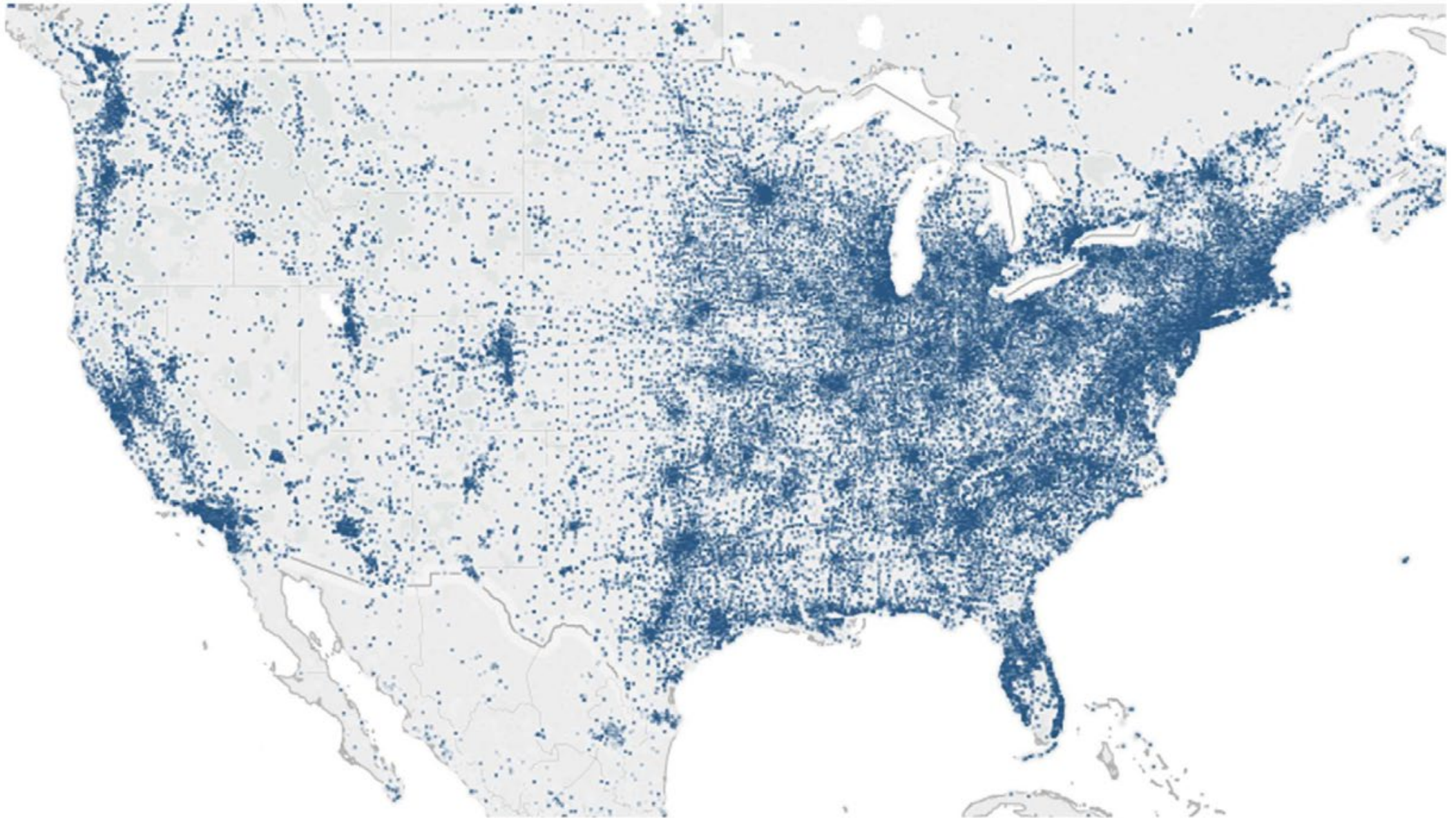
- From 1994, the SEC has required all public domestic companies to submit their filings electronically via EDGAR (Data-Gathering, Analysis, and Retrieval) website
- → centralized disclosure platform
- Before EDGAR, the only central, public source of timely SEC filings was the SEC's reference room in Washington D.C. Studies find evidence of EDGAR reducing awareness costs and acquisition costs,

What is EDGAR Log File Database

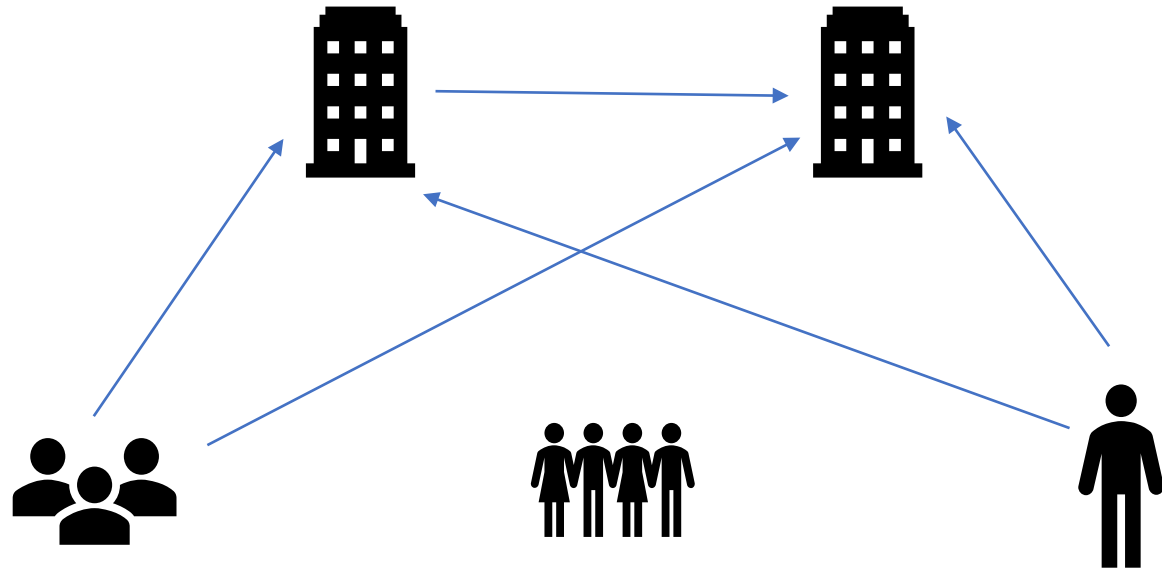
- provide information on [internet search traffic](#) for EDGAR filings through SEC.gov.
- CSV format extracted from log files
- Available by public
 - Daily log files
 - <https://www.sec.gov/about/data/edgar-log-file-data-sets>
- Two datasets:
 - 2003-2017
 - 2020-present: does not contain data from searches originating from SEC IP address ranges.

A sample

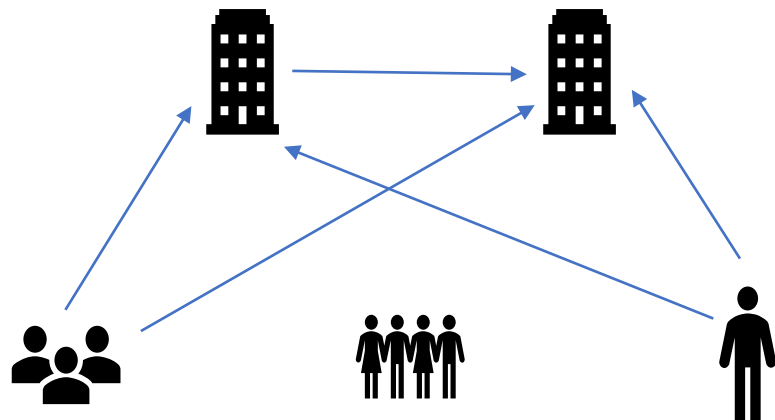
1	ip	date	time	zone	cik	accession	extention	code	size	idx	norefer	noagent	find	crawler	browser
2	129.250.150.jdd	2003/12/31	0:21:03	500	1135444	0001135444	xslF345X02/primary_doc.xml	200	16964	0	1	0	0	0	win
3	129.250.150.jdd	2003/12/31	0:34:00	500	1246350	0000892917	xslF345X02/rcform4thoenes_ex.xml	200	17179	0	1	0	0	0	win
4	129.250.150.jdd	2003/12/31	0:37:44	500	1257612	0000892917	xslF345X02/rcform4erwin_ex.xml	200	25313	0	1	0	0	0	win
5	129.250.150.jdd	2003/12/31	0:41:03	500	1238655	0001238655	xslF345X02/primary_doc.xml	200	15937	0	1	0	0	0	win
6	129.250.150.jdd	2003/12/31	0:44:14	500	1237298	0001237298	xslF345X02/c0001237298d2003122	200	15170	0	1	0	0	0	win
7	67.67.192.cfi	2003/12/31	1:03:46	500	1105101	0001058854	#NAME?	200	2978	1	0	0	1	0	win
8	67.67.192.cfi	2003/12/31	1:03:53	500	1105101	0001058854	gigamedia03-11.txt	200	31611	0	0	0	9	0	win
9	24.128.96.beh	2003/12/31	1:07:45	500	897730	0000950115	.txt	200	40253	0	0	0	1	0	win
10	24.128.96.beh	2003/12/31	1:07:48	500	897730	0001005477	.txt	200	15214	0	0	0	1	0	win
11	24.128.96.beh	2003/12/31	1:07:48	500	897730	0001005477	.txt	200	43825	0	0	0	1	0	win
12	24.128.96.beh	2003/12/31	1:07:49	500	1020520	0000021832	#NAME?	200	29937	1	0	0	1	0	win
13	24.128.96.beh	2003/12/31	1:07:49	500	1020520	0000021832	.txt	200	28167	0	0	0	1	0	win
14	129.250.150.jdd	2003/12/31	1:08:49	500	1187853	0000939057	xslF345X02/primary_doc.xml	200	16969	0	1	0	0	0	win
15	24.128.96.beh	2003/12/31	1:10:12	500	727737	0000898822	#NAME?	200	2349	1	0	0	1	0	win
16	24.128.96.beh	2003/12/31	1:10:12	500	840823	0001164150	.txt	200	75126	0	0	0	1	0	win
17	24.128.96.beh	2003/12/31	1:10:12	500	840823	0001164150	.txt	200	170907	0	0	0	1	0	win
18	205.138.214.ech	2003/12/31	1:10:17	500	2145	0000950123	.hdr.sgml	404	3451	0	1	0	0	1	
19	205.138.214.ech	2003/12/31	1:10:21	500	852015	0000950136	.hdr.sgml	404	3451	0	1	0	0	1	
20	24.128.96.beh	2003/12/31	1:10:21	500	727737	0000927356	.txt	200	31261	0	0	0	1	0	win
21	205.138.214.ech	2003/12/31	1:10:22	500	8177	0001209191	.hdr.sgml	200	962	0	1	0	0	1	
22	24.128.96.beh	2003/12/31	1:10:23	500	727737	0000928385	#NAME?	200	95323	1	0	0	1	0	win
23	24.128.96.beh	2003/12/31	1:16:41	500	1056465	0000950137	#NAME?	200	2540	1	0	0	1	0	win
24	24.128.96.beh	2003/12/31	1:16:41	500	1056465	0000950137	.txt	200	36062	0	0	0	1	0	win
25	24.128.96.beh	2003/12/31	1:16:42	500	1056465	0000950137	#NAME?	200	2305	1	0	0	1	0	win



- With this database, what research questions can you think of?

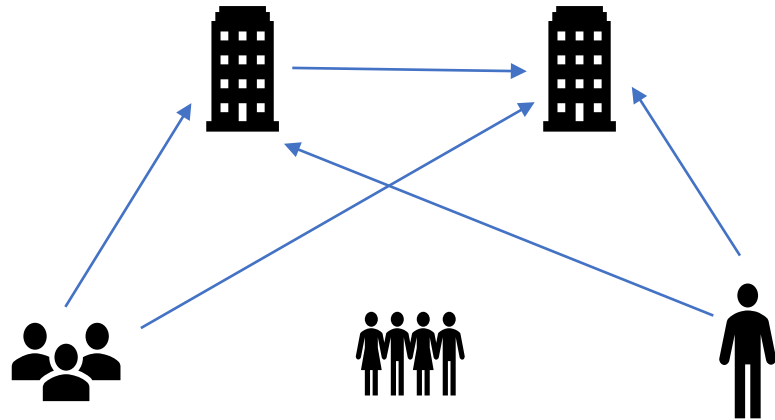


- With this database, what research questions can you think of?



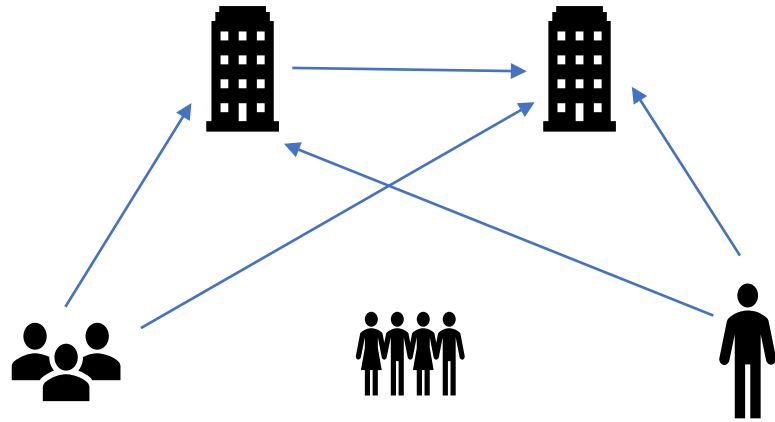
- Who?
 - Investor: ...
 - Analyst: Gibbons et al. 2021
 - mutual fund: Iliev et al. 2021
 - hedge fund: Crane et al. 2023
 - Firms: Bernard, 2020
 - employee/customer/stakeholder/insider/...

- With this database, what research questions can you think of?



- What?
 - 10-K/10-Q
 - Insider trading
 - ...
- Determinants:
 - Performance: Drake, 2015
 - Advertisement: Focke et al. 2020
 - Notification: Chapman, 2018
 - dark pool: Brogaard and Pan, 2022
 - news/...
- Consequences:
 - stock performance: Drake et al, 2020
 - Efficiency
 - Portfolio
 - Investment: Iliev et al, 2021
 - payout...

- With this database, what research questions can you think of?



- Why?
 - Information acquisition: Brogaard and Pan, 2022; Crane et al. 2023...
 - Attention: Drake et al. 2017
 - Monitor: Iliev et al. 2021
 - ...
- Where?
- When?

The selected two papers

- What if the most simple ideas seem to be explored?
- → Higher order information
- → Inspired by real-life story

The wisdom of crowds and the
market's response to earnings news:
evidence using the geographic dispersion of investors

Jason V. Chen, JAE, 2023

Motivation

- Investors possess diverse information → market efficiency
 - → **wisdom of the crowds**
- Some theoretical studies suggest that information diversity among investors can result in price drift
 - (Allen et al., 2006; Hong and Stein, 2007; Banerjee et al., 2009).
- How to measure the investor diversity?
 - Dispersion in investors' geographic locations

Research Question

- General RQ: Does investors' information diversity influence the market efficiency towards firms?
- Specific RQ: Does investors' geographic diversity influence an effective price response to earnings news?

Contribution #1

- Information diversity
- Previous papers: the reader diversity of financial news (Blankespoor, 2019) or the occupational diversity of community residents (Brown, 2008).
- This paper is the first to use geographic location diversity to measure information diversity and demonstrates the feasibility of this measurement approach.

Contribution #2

- The effects of location on investors' information
- Previous papers: investors' distance from a firm's headquarters, which is a first-moment construct of locations
 - (Chi and Shanthikumar, 2017; Coval and Moskowitz, 2001; Grinblatt and Keloharju, 2001; Feng and Seasholes, 2004).
- This paper, considers using dispersion in a firm's investors' locations, a second-moment construct, also has implications for the market's response to earnings news

Contribution #3

- Potential applications for EDGAR data.
- This paper deepens the understanding of EDGAR users by highlighting variations in their locations, which future studies can exploit for other research questions.

Data and variable construction

- Location of the request for filings to the EDGAR filing system
 - Estimate the latitudinal and longitudinal location provided by MaxMind
 - Within the US, precise to postal ZIP code
- 2006.1.1 ~ 2016.3.31
- 16.18 mil requests
 - Remove automated downloads
- Validation:
 - 78.01% originated from the US. 14.89% from New York, most of them come from Manhattan
 - 67% from the state/country mentioned on Business Section and MD&A

Data and variable construction

$$GeoDisp_{i,t} = \log(\sigma_{latitudes}^2 + \sigma_{longitude}^2)$$

- geographic dispersion of the requests for the firm's filings over the year prior to the quarterly earnings announcement date t (trading days $t-255$ to $t-5$, where t is the date of the announcement)
- Control
 - Variation in overall investor interest and attention
 - Local bias to the headquarters
 - Geographic makeup of the firm
 - Other characteristics of EA associated with the response and diversity

Empirical design

- The overall magnitude of the price response during the announcement
- The earnings response coefficient
- Post-earnings announcement drift
- Intra-period timeliness

Empirical results

$$Abs\ ARet_{i,t,a,b} = \beta_0 + \beta_1(High)\ GeoDisp_{i,t} + \beta_i \Sigma Controls + FixedEffects + \varepsilon,$$

Panel A - Magnitude of the Announcement Period Price Response			
Variable	Predict	Full Sample	PSM Sample
		(1)	(2)
		<i>Abs ARet_{i,t-1,1}</i>	<i>Abs ARet_{i,t-1,1}</i>
<i>GeoDisp_{i,t}</i>	+	0.166*** (5.10)	
<i>High GeoDisp_{i,t}</i>	+		0.225*** (3.25)
<i>Size_{i,t}</i>		-0.825*** (-24.59)	-0.844*** (-24.52)
<i>FirmLocations_{i,t}</i>		0.009 (1.61)	0.012* (1.89)
<i>Book-to-Market_{i,t}</i>		-0.178 (-1.47)	-0.442*** (-3.80)
<i>Analysts_{i,t}</i>		0.729*** (16.59)	0.669*** (14.70)
<i>InstOwn_{i,t}</i>		1.408*** (10.67)	1.49*** (9.78)
<i>Requests OL_{i,t}</i>		-0.039 (-0.96)	-0.085** (-2.08)
<i>Requests EA_{i,t}</i>		0.575*** (10.91)	0.695*** (12.47)
<i>Requests Fin EA_{i,t}</i>		0.106*** (3.16)	0.122*** (3.80)

$$ARet_{i,t,a,b} = \beta_0 + \beta_1 (High) GeoDisp_{i,t} + \beta_2 UEarnings_{i,t} + \beta_3 UEarnings_{i,t} \times (High) GeoDisp_{i,t} \\ + \beta_i \Sigma Controls + \beta_i \Sigma UEarnings_{i,t} * Controls + FixedEffects + \epsilon,$$

Panel B - Earnings Response Coefficient			
Variable	Predict	Full Sample	PSM Sample
		(1)	(2)
		$ARet_{i,t,-1,1}$	$ARet_{i,t,-1,1}$
$UEarnings_{i,t}$		1.054*** (10.92)	1.144*** (9.54)
$GeoDisp_{i,t}$		0.042 (1.42)	
$UEarnings_{i,t} \times GeoDisp_{i,t}$	+	0.105*** (10.18)	
$High\ GeoDisp_{i,t}$			0.068 (0.82)
$UEarnings_{i,t} \times High\ GeoDisp_{i,t}$	+		0.192*** (8.50)
$Size_{i,t}$		0.049 (1.10)	0.016 (0.32)
$UEarnings_{i,t} \times Size_{i,t}$		-0.140*** (-11.15)	-0.168*** (-11.18)
$FirmLocations_{i,t}$		0.009** (2.16)	0.012** (2.10)
$UEarnings_{i,t} \times FirmLocations_{i,t}$		-0.002 (-0.71)	-0.002 (-0.82)
$Book-to-Market_{i,t}$		0.682*** (7.86)	0.593*** (5.08)
$UEarnings_{i,t} \times Book-to-Market_{i,t}$		-0.011 (-0.42)	-0.082** (-2.48)

Empirical results

Relation between PEAD and the geographic dispersion of requests.

Panel A - Abnormal Returns Over the Month Following Announcements [$t+2$ to $t+5$, $t+10$, $t+15$, and $t+20$]									
Variable	Predict	Full Sample				PSM Sample			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		$ARet_{i,t,2,5}$	$ARet_{i,t,2,10}$	$ARet_{i,t,2,15}$	$ARet_{i,t,2,20}$	$ARet_{i,t,2,5}$	$ARet_{i,t,2,10}$	$ARet_{i,t,2,15}$	$ARet_{i,t,2,20}$
$UEarnings_{i,t}$ (a)		0.197*** (4.15)	0.170** (2.32)	0.165** (2.06)	0.353*** (3.47)	0.156** (2.58)	0.105 (1.22)	0.045 (0.42)	0.254* (1.82)
$GeoDisp_{i,t}$		-0.005 (-0.23)	-0.020 (-0.52)	-0.035 (-0.65)	-0.004 (-0.05)				
$UEarnings_{i,t} \times GeoDisp_{i,t}$ (b)	-	-0.006 (-0.98)	-0.023** (-2.69)	-0.025** (-2.21)	-0.018 (-1.48)				
$High\ GeoDisp_{i,t}$						-0.021 (-0.52)	-0.074 (-1.20)	-0.162 (-1.60)	-0.115 (-0.90)
$UEarnings_{i,t} \times High\ GeoDisp_{i,t}$ (c)	-					-0.004 (-0.33)	-0.042** (-2.19)	-0.052** (-2.38)	-0.020 (-0.80)
P-value from F-test (H0)									
(a) + (b) * [$75th - 25th\ GeoDisp_{i,t} \geq 0$]		1.000	0.945	0.920	0.999				
(a) + (c) ≥ 0						0.990	0.767	0.475	0.961
$Size_{i,t}$		0.032 (1.26)	0.009 (0.18)	-0.003 (-0.05)	-0.048 (-0.57)	0.037 (1.33)	0.009 (0.19)	-0.017 (-0.24)	-0.071 (-0.81)
$UEarnings_{i,t} \times Size_{i,t}$		-0.025*** (-3.81)	-0.033** (-3.14)	-0.034** (-2.68)	-0.037*** (-2.71)	-0.022*** (-3.00)	-0.029** (-2.39)	-0.026 (-1.56)	-0.032* (-1.90)

Empirical results

- IPT measures how quickly price moves using an area under the curve approach

Inter-period timeliness and the geographic dispersion of requests.

Variable	Predict	Full Sample				PSM Sample			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		$IPT_{i,t,5}$	$IPT_{i,t,10}$	$IPT_{i,t,15}$	$IPT_{i,t,20}$	$IPT_{i,t,5}$	$IPT_{i,t,10}$	$IPT_{i,t,15}$	$IPT_{i,t,20}$
$GeoDisp_{i,t}$	+	0.047*** (3.79)	0.110*** (5.74)	0.066 (1.55)	0.173*** (3.76)				
$High\ GeoDisp_{i,t}$	+					0.076** (2.41)	0.125** (2.12)	0.117 (1.08)	0.222* (1.71)
$Size_{i,t}$		0.028*** (3.13)	0.031* (1.74)	-0.053 (-1.53)	-0.034 (-0.67)	0.029** (2.66)	0.048** (2.02)	-0.058 (-1.60)	-0.026 (-0.40)
$FirmLocations_{i,t}$		0.002 (0.81)	0.006 (1.34)	0.004 (0.64)	0.007 (0.78)	0.001 (0.62)	0.003 (0.72)	0.002 (0.29)	0.010 (0.93)
$Book-to-Market_{i,t}$		0.028 (1.05)	-0.057 (-1.26)	-0.161** (-2.18)	0.026 (0.34)	0.008 (0.22)	-0.131** (-2.31)	-0.262*** (-3.10)	-0.190* (-1.83)

Conclusion

- (1) greater overall magnitude price response during the announcement period, (2) higher ERC, (3) lower PEAD, and (4) higher IPT associated with greater geographic dispersion of requests are consistent with theories that suggest that a greater diversity of information across investors leads to a relatively more efficient price response to earnings news.

IQ from IP: Simplifying search in portfolio choice

Huaizhi Chen, Lauren Cohen, Umit Gurun, Dong Lou, Christopher Malloy

JFE, 2020

IQ from IP: Simplifying search in portfolio choice

- Motivation

- In light of the decreasing cost of creating, processing, and transmitting info, how does an investor reduce the dimensionality and know which subset of signals have the potential to be informative?
 - How investors approach this foundational problem remains a black box.
 - This paper provides micro-level foundation in the search process in delegated portfolio management.

- Setting

- Why fund?
 - Managers have a comparative advantage relative to peers → structure on info acquisition
- Why insider trades?
 - Potential attractive candidate for comparative advantage signals for mutual fund managers
 - e.g., AOK Inc. Independent board member's insider trade, fund manager followed

Research question

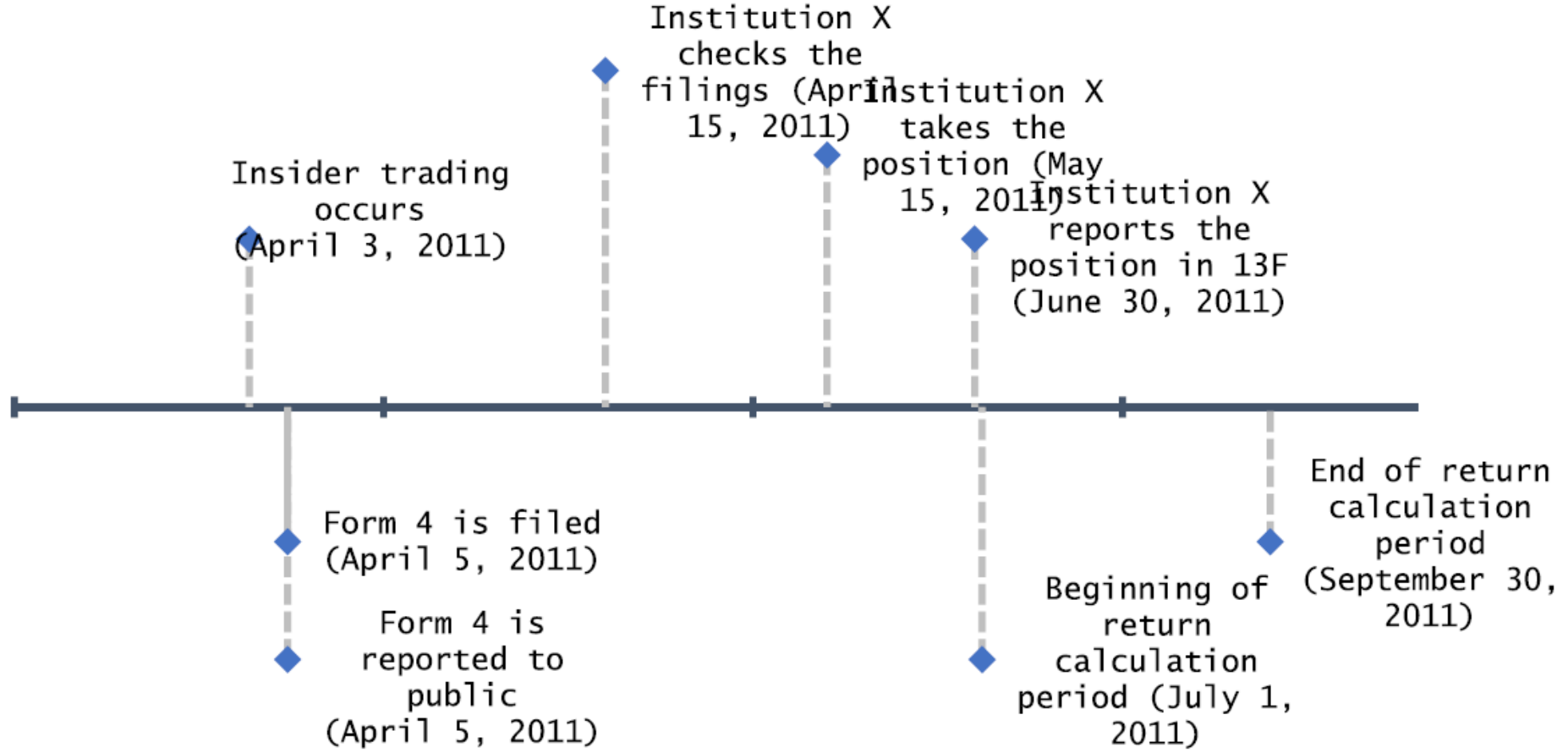
- Whether fund managers track firm insider information?
- How tracked trades have predictability for future firm operation and returns?
- How corresponding portfolio strategy predict funds' returns?

Contribution

- Investment performance of mutual fund managers
 - Prior research: stock-picking by distance, network, ...
 - This paper explores to what extent mutual funds actively investigate the insider trades within their portfolio
- Behavior of corporate insider trading
 - Prior research: firm-level insider trades
 - This paper explores how individual-level insider trades provide information
- Gradual information diffusion and limited attention
 - This paper investigates a novel channel of the information diffusion

Data

- Matching IP address to 13-F organizations
 - Decipher the IP addresses' hidden octets
 - Hand-match names of the 13-F organizations to the IP address
 - Scrape the insider trading filings from the SEC website and merge by code
- Individual
 - Match the institutional identifier from Ancerno database (managercode) to the institution identifier (Mgrno)
 - BoardEX: Information on corporate insiders: academic qualification; current and past job; membership
- Morningstar database: fund-level performance; manager information



Empirical design

- Persistence in tracking behavior
- Contemporaneous trading of fund managers & insiders
- Portfolio returns to active insider tracking
- Portfolio returns across firm and insider characteristics
- Mechanism
 - Characteristics of insiders
 - Characteristics of fund managers
 - Source of fund manager insider links

Empirical result

Panel A. Persistence of tracking behavior				Specific insider filings		
	downloads at least one insider trading filing by a given firm					
	Checked firm at $t + 1$			Checked insider at $t + 1$		
Checked firm at t	0.413*** (14.37)	0.259*** (12.20)	0.255*** (11.86)			
Checked insider at t				0.187*** (9.20)	0.142*** (6.58)	0.133*** (6.31)
Portfolio FE	No	Yes	Yes	No	Yes	Yes
Year x Stock FE	No	No	Yes			
Year x Stock x Insider FE				No	No	Yes
Adj. R ²	0.1947	0.2696	0.2833	0.0387	0.0842	0.0958
No.	1338,919	1338,919	1338,919	11,190,087	11,190,087	11,190,087

- Contemporaneous trading of fund managers and corporate insiders

	FundDirection (−1 for sell, 1 for buy)					
CheckedInsiderDirection	0.087*** (4.75)			0.079*** (4.22)		
CheckedInsiderBuy		0.170*** (6.60)			0.163*** (5.91)	
CheckedInsiderSell			−0.095*** (−3.77)			−0.082*** (−3.37)
UncheckedInsiderDirection	0.014*** (3.44)			0.015*** (3.75)		
UncheckedInsiderBuy		0.023*** (3.38)			0.026*** (3.59)	
UncheckedInsiderSell			−0.037*** (−4.03)			−0.040*** (−4.39)
Time FE	Yes	Yes	Yes	No	No	No
Institution FE	Yes	Yes	Yes	No	No	No
Time x Institution FE	No	No	No	Yes	Yes	Yes
Wald test: (prob > F)	0.001	0.001	0.037	0.002	0.001	0.119
Adj. R ²	0.0264	0.0265	0.0264	0.908	0.909	0.907
No. obs.	2469,803	2469,803	2469,803	2469,792	2469,792	2469,792

- Portfolio returns to active insider tracking

Panel A: Tracked Insider Buys	Excess returns	DGTW	4-factor alpha	L% of assets	L 4F alpha	S% of assets	S 4F alpha
1) All positions returns of the average stock held within an institutional portfolio	2.78% (1.92)	0.22% (2.02)	0.34% (2.16)	100%			
2) All positions except checked insider buying all holdings where the institution does not check	2.76% (1.93)	0.22% (1.93)	0.33% (2.14)	98.7%			
3) Checked insider buying vs 2)	1.87% (1.56)	2.02% (1.84)	1.77% (1.57)	1.28%	2.10% (1.84)	98.7%	0.33% (2.14)
4) Checked insider buying and bought vs 2) all stocks of which a fund manager checks the insiders' transaction and also trade shares in the same direction	2.96% (2.02)	2.90% (2.12)	3.00% (2.08)	0.68%	3.33% (2.30)	98.7%	0.33% (2.14)
5) Checked insider buying and bought vs. rest bought	2.97% (2.03)	2.90% (2.13)	3.00% (2.10)	0.68%	3.33% (2.30)	49.8%	0.33% (1.93)
6) Checked insider buying and bought vs. checked and not bought	3.68% (2.13)	2.55% (1.63)	3.25% (1.91)	0.68%	3.33% (2.30)	0.61%	0.08% (0.08)
7) Checked insider buying and bought vs. checked and not bought (zero initial positions)	4.28% (2.39)	2.90% (1.79)	3.93% (2.23)	0.15%	3.80% (2.46)	0.15%	-0.14% (-0.13)
8) Checked insider buying and bought vs. not checked insider buying and bought	3.02% (2.17)	2.97% (2.22)	2.97% (2.09)	0.68%	3.33% (2.30)	2.65%	0.35% (0.95)

- Mechanism of the link

	Match indicator					
Education link	0.328%** (2.64)	0.349%*** (2.76)	0.325%*** (2.75)	0.370%*** (3.85)	0.668%*** (6.20)	0.391%*** (3.91)
Location link	1.42%*** (4.84)	1.40%*** (4.98)	1.31%*** (4.77)	0.955%*** (3.92)	0.501%** (2.15)	0.502%** (2.20)
Lag match indicator			21.9%*** (12.32)	18.9%*** (11.86)	20.04%*** (11.39)	17.91%*** (11.26)
Quarter FE	No	Yes	Yes	Yes	Yes	Yes
Insider education FE	No	No	No	Yes	No	Yes
Manager education FE	No	No	No	Yes	No	Yes
Insider location FE	No	No	No	No	Yes	Yes
Manager location FE	No	No	No	No	Yes	Yes
Adj-R ²	0.0005	0.0141	0.0303	0.0533	0.0455	0.0618
No. obs.	1748,892	1748,892	1748,892	1748,892	1705,546	1705,546

Conclusion

- Using web traffic on the SEC's EDGAR server between 2004 and 2015, we find that mutual fund managers track a very particular subset of firms and insiders.
- Fund manager tracking activity not only remains persistent over time, but also has powerful implications for their portfolio choice.
- The trades they track and choose to act upon significantly outperform those that they track and choose not to trade along with

Summary

- EDGAR log file database is very informative to tell a story.
- We can utilize the geographic features and complement more future studies.
 - New settings
 - Combination with other alternative data
- Increasing hurdle