Data Scientists on Wall Street

Ling Cen[1] Bing Han[2] Yanru Han[3] Chanik Jo[1]

ABFER

[1] CUHK Business School, The Chinese University of Hong Kong

[2] Rotman School of Management, The University of Toronto

[3] Stevens Institute of Technology

Motivation

- The finance industry is experiencing a significant transformation due to the advent of big data and artificial intelligence.
- Traditional data: corporate disclosures (e.g., annual reports and corporate announcements), information intermediaries (e.g., Bloomberg, Reuters, and financial analysts), and macroeconomic news (FRED, BLS, BEA), etc.

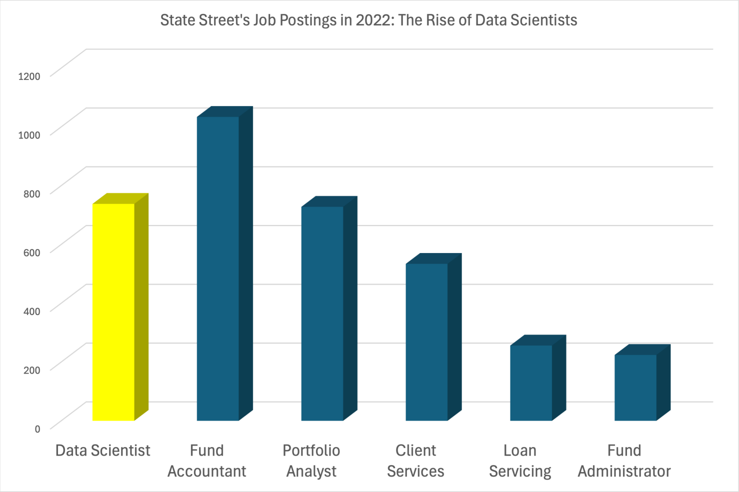
Motivation

- New data: A broader and often less-structured set of information that is typically sourced from the digital and physical footprint of economic activity—e.g., Website scraping with NLP (e.g., customers' rating, social media), job postings (hiring trends), satellite, shipping/cargo, credit card tracking (transaction patterns), wealth, geolocation, etc.

- Extracting valuable information from big data requires the expertise of data scientists (Huang et al. 2022)
- Data scientists collect, maintain, and analyze unstructured alternative data (JP Morgan, 2017)

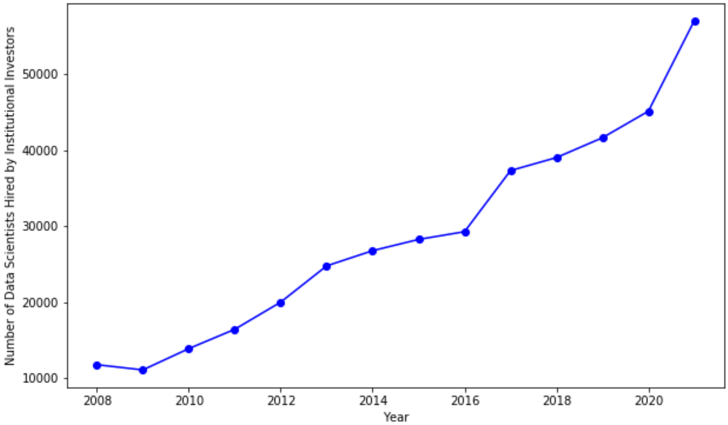
The Battle for Data Scientists is Reshaping Wall Street

- In 2022, State Street posted over 700 job openings for data scientists, surpassing the number of listings for traditional portfolio analyst roles.



The Battle for Data Scientists is Reshaping Wall Street

- The total number of data scientists employed by 13F institutional investors has more than quadrupled (i.e., 11,799 to 57,050) from 2008 to 2021.
- Why do institutional investors aggressively recruit data scientists?



This Paper Asks:

- Why do financial institutions aggressively recruit data scientists?
  - ✓ Do data scientists help financial institutions generate abnormal returns, i.e., alphas?
- When do institutional investors gain the most advantage from data scientists?
- How do financial institutions maintain the advantages brought by data scientists?
- Labor market competition?
  - ✓ Are they catching up with leading competitors? How do they know?
- Implication for overall financial markets: the stock price informativeness?

This Paper Asks:

- Why do financial institutions aggressively recruit data scientists?
  - ✓ Do data scientists help financial institutions generate abnormal returns, i.e., alphas?
- When do institutional investors gain the most advantage from data scientists?
- How do financial institutions maintain the advantages brought by data scientists?
- Labor market competition?
  - ✓ Are they catching up with leading competitors? How do they know?
- Implication for overall financial markets: the stock price informativeness?

This Paper Asks:

- Why do financial institutions aggressively recruit data scientists?
  - ✓ Do data scientists help financial institutions generate abnormal returns, i.e., alphas?
- When do institutional investors gain the most advantage from data scientists?
- How do financial institutions maintain the advantages brought by data scientists?
- Labor market competition?
  - ✓ Are they catching up with leading competitors? How do they know?
- Implication for overall financial markets: the stock price informativeness?

This Paper Asks:

- Why do financial institutions aggressively recruit data scientists?
  - ✓ Do data scientists help financial institutions generate abnormal returns, i.e., alphas?
- When do institutional investors gain the most advantage from data scientists?
- How do financial institutions maintain the advantages brought by data scientists?
- Labor market competition?
  - ✓ Are they catching up with leading competitors? How do they know?
- Implication for overall financial markets: the stock price informativeness?

This Paper Asks:

- Why do financial institutions aggressively recruit data scientists?
  - ✓ Do data scientists help financial institutions generate abnormal returns, i.e., alphas?
- When do institutional investors gain the most advantage from data scientists?
- How do financial institutions maintain the advantages brought by data scientists?
- Labor market competition?
  - ✓ Are they catching up with leading competitors? How do they know?
- Implication for overall financial markets: the stock price informativeness?

This Paper Find:

- Why do financial institutions aggressively recruit data scientists?

$\Rightarrow$ Institutional investors who hire more data scientists achieve higher trading profitability

$\Rightarrow$ Each additional data scientist hired by an investor is associated with a 0.004 percentage point higher CAPM alpha per quarter (versus 0.031% average trading profitability in our sample – i.e., 13% relative to the mean).

- When do institutional investors gain the most advantage from data scientists?

$\Rightarrow$ The positive impact of hiring data scientists on trading profitability is amplified when trading stocks with a high data scientist concentration.

- How do financial institutions maintain the advantages brought by data scientists?

$\Rightarrow$ Institutional investors strategically adjust their portfolio allocation and recruitment decisions to maximize the benefits provided by data scientists $\Rightarrow$ tilt portfolio towards stocks with a high data scientist concentration

$\Rightarrow$ Institutional investors' portfolios become more concentrated

This Paper Find:

- Why do financial institutions aggressively recruit data scientists?

⇒ Institutional investors who hire more data scientists achieve higher trading profitability

⇒ Each additional data scientist hired by an investor is associated with a 0.004 percentage point higher CAPM alpha per quarter (versus 0.031% average trading profitability in our sample – i.e., 13% relative to the mean).

- When do institutional investors gain the most advantage from data scientists?

⇒ The positive impact of hiring data scientists on trading profitability is amplified when trading stocks with a high data scientist concentration.

- How do financial institutions maintain the advantages brought by data scientists?

⇒ Institutional investors strategically adjust their portfolio allocation and recruitment decisions to maximize the benefits provided by data scientists ⇒ tilt portfolio towards stocks with a high data scientist concentration

⇒ Institutional investors' portfolios become more concentrated

This Paper Find:

- Why do financial institutions aggressively recruit data scientists?

$\Rightarrow$ Institutional investors who hire more data scientists achieve higher trading profitability

$\Rightarrow$ Each additional data scientist hired by an investor is associated with a 0.004 percentage point higher CAPM alpha per quarter (versus 0.031% average trading profitability in our sample – i.e., 13% relative to the mean).

- When do institutional investors gain the most advantage from data scientists?

$\Rightarrow$ The positive impact of hiring data scientists on trading profitability is amplified when trading stocks with a high data scientist concentration.

- How do financial institutions maintain the advantages brought by data scientists?

$\Rightarrow$ Institutional investors strategically adjust their portfolio allocation and recruitment decisions to maximize the benefits provided by data scientists $\Rightarrow$ tilt portfolio towards stocks with a high data scientist concentration

$\Rightarrow$ Institutional investors' portfolios become more concentrated

This Paper Find (cont'd):

- Labor market competition?

⇒ Investors keep up with their key competitors with the most data scientists

⇒ Stronger when you hire data scientists from key competitors (information)

⇒ Placebo – when you don't care when the gap between non-competitors who have the most data scientists

- Implication for overall financial markets: the stock price informativeness?

⇒ The concentration of data scientists among a few investors diminishes price informativeness.

This Paper Find (cont'd):

- Labor market competition?

$\Rightarrow$ Investors keep up with their key competitors with the most data scientists

$\Rightarrow$ Stronger when you hire data scientists from key competitors (information)

$\Rightarrow$ Placebo – when you don't care when the gap between non-competitors who have the most data scientists

- Implication for overall financial markets: the stock price informativeness?

$\Rightarrow$ The concentration of data scientists among a few investors diminishes price informativeness.

The Role of Data Scientists in the Big Data Era

- Existing studies on the role of big data in the financial market
  - ✓ Farboodi et al. (2022) develop a quantitative measure of data usage by investors. Farboodi et al. (2019), Farboodi and Veldkamp (2022), and Veldkamp (2023) provide a theoretical framework and a set of tools to value data as an asset.
  - ✓ Various implications: Price informativeness (Dugast and Foucault, 2018), disciplinary effect of corporate managers (Zhu, 2019), endogenous data skill acquisition (Huang et al., 2022), market efficiency (Martin and Nagel, 2022), economics of knowledge production (Abis and Veldkamp, 2024), capital allocation (Dugast and Foucault, 2024), forecast horizon (Dessaint et al., 2024), a networked asset (Bian et al., 2025), forecast accuracy (Chi et al., 2025), and information asymmetry among investors (Katona et al., 2025)

- How We Differ:
  - ✓ Unlike data that can be shared with a positive externality, we focus on data scientists who cannot be employed by multiple institutional investors at the same time.
  - ✓ Examine how competition for data talent and its concentration affect investment performance and price informativeness.

Contributions

- Institutional investors' use of AI and big data in asset management: Bonelli and Foucault (2023), Sheng et al. (2024), Abis et al. (2025), and Bonelli (2025), among others.

$\Rightarrow$ First paper to examine data scientists hiring by asset management firms, demonstrating their alpha generation, strategic adjustments in portfolios and hiring, and concentration's role in reducing price informativeness via competition.

- Our paper also contributes to the growing literature on the labor market of financial workers (Oyer, 2008, Ellul et al. 2019)

- We identify a prevailing trend of recruiting data scientists among institutional investors and show that data scientists causally help financial institutions generate abnormal returns.

Data: Quarterly 13F institutional investors' holdings data

- LSEG Reuters Global Ownership Database
- The dataset reports holdings of any institutions with assets under management of $100 million or more, covering a wide range of financial institutions.
- Institution type (hedge fund, investment advisor, bank, insurer, PE, VC, sovereign wealth fund, etc), detailed location (address and zip code), etc.

Data: Institutional Investors' Employment of Data Scientists

- Detailed resumes of employees from Revelio Lab database.
- Revelio gathers career history data from various unstructured public online sources, including LinkedIn (= subset).
- Data scientists are identified based on the occupation codes (ONET codes) associated with their positions.
- Merging the Revelio data with institutional holding data from the LSEG Reuters Global Ownership database
  - ✓ 124,947 unique data scientists employed by 1,957 institutional investors.
- The data starts from the 2000s, but as in previous studies (Cai et al., 2022; Liang et al., 2023), we use the sample from 2008.
  - ✓ 11,799 (57,050) data scientists employed by institutions in 2008 (2021)

Data: Institutional Investors' Employment of Data Scientists

- Data collection: gathering and organizing data
- Data analytics: analyzing data for making business decisions
- Data maintenance: storing and protecting data.

ONET Code	Occupation	Data Scientists Category
15-2051.00	Data Scientists	Data Analytics
15-2041.00	Statisticians	Data Analytics
15-1299.06	Digital Forensics Analysts	Data Analytics
15-2051.01	Business Intelligence Analysts	Data Analytics
15-2051.02	Clinical Data Managers	Data Analytics
15-1242.00	Database Administrators	Data Collection
15-1243.00	Database Architects	Data Collection
15-1212.00	Information Security Analysts	Data Maintenance
15-1243.01	Data Warehousing Specialists	Data Maintenance
15-1299.05	Information Security Engineers	Data Maintenance
15-1299.04	Penetration Testers	Data Maintenance

Data: Leading Institutional Investors by Data Scientist Employment

Investor	NumDS	NumAnalysis	NumCollect	NumMaintain
Morgan Stanley & Co. LLC	4311	2746	259	1306
Credit Suisse Asset Management, LLC (US)	3288	1877	245	1166
Goldman Sachs & Company, Inc.	3076	2213	95	768
Liberty Mutual Insurance Group	2353	1535	126	692
Blackrock Alternatives Management, LLC	1820	1484	52	284
Fidelity National Financial Inc.	685	468	48	169
Bank of the West	617	293	19	305
LPL Financial LLC	389	203	25	161
TD Securities, Inc.	352	282	7	63
Brown Brothers Harriman & Company	325	186	29	110
Wells Fargo	309	242	10	57
Kemper Corporation	252	153	17	82
Susquehanna International Group, LLP	237	115	33	89
First Bancorp, Inc	230	87	27	116
Protective Life Corporation	205	148	10	47
PIMCO (US)	196	102	17	77
Barclays Capital Inc.	179	129	13	37

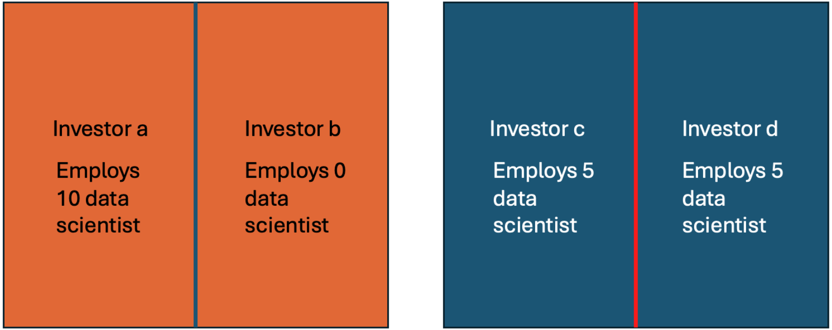
Data: Measuring Data Scientists Coverage at Stock Level

- How many data scientists are associated with stock j?
- For a single stock, a data scientist of a major shareholder is not equivalent to a data scientist of a minor shareholder

$$\textit{DS Coverage}_{j,t} = \sum_i \textit{NumDS}_{i,t} \times \frac{\textit{Shares Held}_{i,j,t}}{\textit{Shares Outstanding}_{j,t}}, \tag{1}$$

- By constructing this measure, we implicitly assume that data scientists of major shareholders exert a greater influence on stock prices relative to those of minor shareholders or those of financial institutions that have no position in this stock.

Data: Measuring Data Scientists Concentration at Stock Level



Company A

Company B

$$DS\ HHI_{j,t} = \sum_i \left(\frac{NumDS_{i,t} \times \frac{Shares\ Held_{i,j,t}}{Shares\ Outstanding_{j,t}}}{DS\ Coverage_{j,t}} \right)^2, \quad (2)$$

Data: Institutional Characteristics and Data Scientist Hiring

Panel A. Institutional Investor Characteristics and the Hiring of Data Scientists

	(1)	(2)	(3)
		Num DS	
LogTNA	-0.038	-0.075	-0.071
	(0.122)	(0.127)	(0.139)
Num NonDS Employee	0.041***	0.041***	0.042***
	(0.000)	(0.001)	(0.001)
Log Number of SIC	0.796***	0.852***	0.744**
	(0.172)	(0.254)	(0.313)
Turnover	2.771***	2.927**	3.672***
	(0.979)	(1.264)	(1.293)
HedgeFund	0.859**	0.970***	
	(0.397)	(0.311)	
PensionFund	-18.391***	-18.293***	
	(1.821)	(4.883)	
Bank	-3.099***	-2.961***	
	(0.935)	(0.711)	
NumProgram	0.732***	0.599***	0.618***
	(0.194)	(0.199)	(0.200)
Institution Type × Time FEs	×	×	✓
Time FEs	×	✓	×
Obs.	49,617	49,617	49,601
Adj. R2	0.885	0.885	0.888

Why Do Financial Institutions Aggressively Recruit Data Scientists?

- Do data scientists help financial institutions generate abnormal returns, i.e., alphas, in portfolio management?
- We estimate the following regression for 13F investor i at quarter t:

$$Alpha_{i,t+1} = \beta_0 + \beta_1 NumDS_{i,t} + Investor\ Controls + Fixed\ Effects, \tag{3}$$

- $Alpha_{i,t+1}$ denotes trading profitability following Kumar et al. (2020) and Bonelli and Foucault (2023)
- CAPM α_{t+1}, for instance, is calculated as:

$$CAPM\ \alpha_{i,t+1} = \sum_j (Weight_{i,j,t} - Weight_{i,j,t-1}) \times CAPM\ \alpha_{j,t+1}, \tag{4}$$

Why Do Financial Institutions Aggressively Recruit Data Scientists?

	(1) CAPM α_{t+1}	(2) FF3 α_{t+1}	(3) FF4 α_{t+1}
NumDS	0.004***	0.002**	0.002***
	(0.001)	(0.001)	(0.001)
LogTNA	0.034***	0.033***	0.036***
	(0.005)	(0.004)	(0.004)
Log Number of Firm	0.002	-0.004	-0.011*
	(0.006)	(0.005)	(0.005)
Turnover	-0.026	-0.002	-0.064**
	(0.028)	(0.028)	(0.027)
FewIndDummy	-0.009	0.005	0.016
	(0.020)	(0.019)	(0.019)
Log Market Cap	-0.000	-0.001	-0.000
	(0.000)	(0.000)	(0.000)
Volume	-0.238***	0.024	-0.142**
	(0.060)	(0.056)	(0.057)
Gross Profit	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)
Investor FEs	✓	✓	✓
Time FEs	✓	✓	✓
Obs.	186,478	186,478	186,478
Adj. R2	0.087	0.091	0.090

Identification: Exogenous Shock to the Local Availability of Data Scientists

- In Table 2, we show that the number of data scientists hired by institutional investors is positively correlated with the number of data-science-related undergraduate programs offered by local universities
- The establishment of new data science undergraduate programs
  - ✓ Largely independent of local institutional investors' actions
  - ✓ The four-year lag between the introduction of new programs and the availability of graduates makes it difficult for institutional investors to time or control the supply, supporting the exogeneity of this shock in our context.
- We count the cumulative number of local data scientist undergraduate programs established four years prior ($t-4$) in the same state as the investor (*NumProgram*).
- We employ a two-stage least squares (2SLS) approach, using *NumProgram* as the instrumental variable for the number of data scientists hired by the investor.

Identification: Exogenous Shock to the Local Availability of Data Scientists

	(1) First Stage	(2) Second Stage	(3) First Stage	(4) Second Stage	(5) First Stage	(6) Second Stage	(7) First Stage	(8) Second Stage
	NumDS	CAPM α_{t+1}	NumDS Analysis	CAPM α_{t+1}	NumDS Collect	CAPM α_{t+1}	NumDS Maintain	CAPM α_{t+1}
$\widehat{\text{NumDS}}$		0.095***						
		(0.019)						
$\widehat{\text{NumDS Analysis}}$				0.182***				
				(0.040)				
$\widehat{\text{NumDS Collect}}$						2.374**		
						(1.044)		
$\widehat{\text{NumDS Maintain}}$								0.233***
								(0.046)
NumProgram	0.327***		0.170***		0.013**		0.133***	
	(0.052)		(0.031)		(0.006)		(0.021)	
LogTNA	0.319***	0.002	0.191***	-0.003	0.015**	-0.005	0.113***	0.006
	(0.054)	(0.009)	(0.030)	(0.010)	(0.006)	(0.023)	(0.020)	(0.008)
Log Number of Firm	0.098	-0.006	0.055	-0.007	0.014	-0.030	0.022	-0.002
	(0.076)	(0.009)	(0.046)	(0.010)	(0.009)	(0.026)	(0.026)	(0.008)
Turnover	0.147	-0.039	0.078	-0.039	-0.004	-0.015	0.078*	-0.043
	(0.100)	(0.030)	(0.055)	(0.030)	(0.012)	(0.039)	(0.040)	(0.030)
FewIndDummy	0.577***	-0.065**	0.263***	-0.059**	0.052**	-0.133*	0.229***	-0.064**
	(0.148)	(0.027)	(0.101)	(0.029)	(0.021)	(0.075)	(0.052)	(0.025)
Log Market Cap	0.003	-0.000	0.002	-0.000	0.000	-0.001	0.000	-0.000
	(0.005)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Volume	-1.830***	-0.041	-0.927***	-0.045	-0.122***	0.075	-0.740***	-0.041
	(0.401)	(0.076)	(0.209)	(0.077)	(0.045)	(0.160)	(0.197)	(0.082)
Gross Profit	-0.000**	0.000	-0.000**	0.000	-0.000*	0.000	-0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Investor FEs	✓	✓	✓	✓	✓	✓	✓	✓
Time FEs	✓	✓	✓	✓	✓	✓	✓	✓
Obs.	186,478	186,478	186,478	186,478	186,478	186,478	186,478	186,478
F-statistic	557.998		419.085		60.290		648.098	

When Do Institutions Gain the Most Advantage from Data Scientists?

- Investors could gain the most advantage from data scientists when coverage is concentrated among fewer investors, as this reduces competition for information and enhances unique insights.

- We use an investor-firm-quarter level sample to examine the relationship between trading profitability of institution *i* and the concentration of data scientists covering the firm *j*:

$$\begin{aligned}
Alpha_{i,j,t+1} = \beta_0 &+ \beta_1 PortfolioWeighted\ NumDS_{i,j,t} \times DS\ HHI_{j,t} \\
&+ \beta_2 DS\ HHI_{j,t} + \beta_3 PortfolioWeighted\ NumDS_{i,j,t} \\
&+ Investor\ Controls + Firm\ Controls + Fixed\ Effects,
\end{aligned} \tag{5}$$

where $PortfolioWeighted\ NumDS_{i,j,t} = weight_{i,j,t} \times NumDS_{i,t}$

When Do Institutions Gain the Most Advantage from Data Scientists?

	(1) CAPM $\alpha_{i,j,t+1}$	(2) FF3 $\alpha_{i,j,t+1}$	(3) FF4 α_{t+1}
Portfolio-Weighted NumDS × DS HHI	0.017	0.017**	0.018***
	(0.013)	(0.007)	(0.006)
DS HHI	-0.103***	-0.104***	-0.114***
	(0.010)	(0.010)	(0.010)
Portfolio-Weighted NumDS	-0.009	-0.006	-0.008
	(0.012)	(0.007)	(0.006)
LogTNA	-0.004	0.005	0.001
	(0.004)	(0.004)	(0.004)
Log Number of Firm	-0.007	-0.007	0.005
	(0.006)	(0.005)	(0.005)
Turnover	-0.100**	-0.025	-0.098**
	(0.042)	(0.040)	(0.042)
FewIndDummy	-1.722**	-0.616	-0.275
	(0.751)	(0.704)	(0.667)
LogAsset	0.141***	0.118***	0.121***
	(0.007)	(0.006)	(0.006)
Tobin Q	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
ROA	-0.005***	-0.004***	-0.005***
	(0.001)	(0.001)	(0.001)
LogAge	0.014*	-0.041***	-0.019**
	(0.008)	(0.008)	(0.008)
Investor FEs	✓	✓	✓
Firm FEs	✓	✓	✓
Time FEs	✓	✓	✓
Obs.	30,640,198	30,638,578	30,636,392
Adj. R2	0.008	0.007	0.008

How Do Financial Institutions Maintain Advantages Brought by Data Scientists?

- Do investors strategically tailor their portfolios to maximize the advantages provided by their data scientists?
  - ✓ If much of the data is unique to each firm—such as proprietary datasets, disclosures, and alternative data—Investors with data scientists would concentrate their holdings on a smaller set of stocks.
  - ✓ Given the previous result, investors with data scientists may strategically hold portfolios with stocks that have higher data scientist concentration.

How Do Institutions Adjust Portfolios to Benefit from Their Data Scientists?

- Holding $\text{HHI}_{i,t+1} = \sum_j \mathit{Weight}^2_{i,j,t}$
  - ✓ Concentration of investor *i*'s portfolio
- Portfolio DS $\text{HHI}_{i,t+1} = \sum_j \mathit{Weight}_{i,j,t} \times \mathit{DS\ HHI}_{j,t}$
  - ✓ Average data scientist concentration of investor *i*'s portfolio

How Do Institutions Adjust Portfolios to Benefit from Their Data Scientists?

- Investors who hire more data scientists tend to hold more concentrated portfolios and strategically focus on stocks with higher data scientist concentration.

	(1) Holding HHI_{t+1}	(2) Portfolio DS HHI_{t+1}
NumDS	0.004***	0.004***
	(0.001)	(0.001)
LogTNA	-0.690***	0.627***
	(0.108)	(0.122)
Log Number of Firm	-4.861***	-0.588***
	(0.285)	(0.092)
Turnover	0.839	-1.217***
	(0.577)	(0.368)
FewIndDummy	15.607***	1.513**
	(0.908)	(0.584)
Log Market Cap	-0.019*	-0.090***
	(0.010)	(0.015)
Volume	0.496	0.457
	(1.476)	(3.472)
Gross Profit	-0.005**	0.001
	(0.003)	(0.002)
Investor FEs	✓	✓
Time FEs	✓	✓
Obs.	186,478	186,478
Adj. R2	0.840	0.621

Do Institutions Keep Up with the Talent Race for Data Scientists?

- Investors may keep up with their rivals' hiring activities and seek to close gaps in data scientist staffing

- $\textit{DS Diff}_{i,t} = \sum_j [(\textit{NumDS}_{j,t} - \textit{NumDS}_{i,t}) \times \textit{DS HHI}_{j,t} \times \textit{Weight}_{i,j,t}]$ where the sum is over all stocks j held by institution i, and $\textit{NumDS}_{j,t}$ is the number of data scientists employed by the institution with the most data scientists holding stock j (the leading competitor).

- $\textit{Overlap NumDS Diff}_{i,t} = \sum_j [(\textit{NumDS}_{j,t} - \textit{NumDS}_{i,t}) \times \textit{DS HHI}_{j,t} \times \textit{Weight}_{i,j,t} \times \textit{Num Overlap DS}_{i,j,t}]$ where $\textit{NumOverlapDS}_{i,j,t}$ is the number of data scientists hired by i who previously worked for the leading competitor holding stock j.

Hiring Data Scientists to Keep Up with Competitors

	(1)	(2)	(3)	(4)
	Δ NumDS$_{t+2}$			
NumDS Diff	0.009**	0.010**	0.010**	0.010**
	(0.004)	(0.004)	(0.004)	(0.004)
Overlap NumDS Diff			0.075**	
			(0.037)	
Placebo NumDS Diff				0.001
				(0.010)
LogTNA		-0.001	-0.001	-0.001
		(0.004)	(0.004)	(0.004)
Log Number of Firm		0.028**	0.028**	0.028**
		(0.013)	(0.013)	(0.013)
Turnover		0.022	0.021	0.022
		(0.027)	(0.027)	(0.027)
FewIndDummy		0.034	0.036	0.035
		(0.052)	(0.052)	(0.052)
Log Market Cap		0.003**	0.003**	0.003**
		(0.001)	(0.001)	(0.001)
Volume		-0.190	-0.184	-0.189
		(0.129)	(0.129)	(0.129)
Gross Profit		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Investor FEs	✓	✓	✓	✓
Time FEs	✓	✓	✓	✓
Obs.	44,196	44,196	44,196	44,196
Adj. R2	0.4099	0.4101	0.4110	0.4101

Implication for price informativeness

- While data scientists on Wall Street provide their employers with information advantage and benefit their trading profitability, it may not enhance stock market information efficiency.
- When valuable information is concentrated among a few investors, the dominant investors lack incentives to trade swiftly on their insights, and the broader market lacks access to this information
  - ✓ Delays the incorporation of new information into stock prices, reducing the price informativeness of the affected stocks.

Implication for price informativeness

- Information theories
  - ✓ Kyle (1985): Information monopoly $\rightarrow$ price incorporates information gradually $\rightarrow$ Price efficiency $\downarrow$
  - ✓ Holden and Subrahmanyam (1992): Multiple informed investors trade aggressively $\rightarrow$ Price efficiency $\uparrow$

Hypothesis

- If the number of data scientists can capture information flow that investors have,
  - ✓ The high concentration of data scientists $\approx$ information monopoly $\rightarrow$ Price efficiency $\downarrow$
  - ✓ The low concentration of data scientists $\approx$ fierce information competition $\rightarrow$ Price efficiency $\uparrow$
- Otherwise, there is no reason to believe that stock price informativeness has anything to do with the concentration of data scientists

Implication for price informativeness

- Following Bai et al. (2016), Kacperczyk et al. (2021), and Xiong et al. (2024), we measure price informativeness by examining how well current market prices predict future cash flows:

$$\begin{aligned} Earnings_{j,t+h} = &\beta_0 + \beta_1 LogMVA_{j,t} + \beta_2 LogMVA_{j,t} \times DS\ HHI_{j,t} + \beta_3 DS\ HHI_{j,t} \\ &+ \beta_4 LogMVA_{j,t} \times Log\ DS\ Coverage_{j,t} + \beta_5 Log\ DS\ Coverage_{j,t} \\ &+ LogMVA_{j,t} \times Firm\ Controls + Firm\ Controls + Fixed\ Effects \end{aligned} \tag{6}$$

Implication for price informativeness

	(1) Earnings$_{t+1}$	(2) Earnings$_{t+3}$
LogMVA	0.039***	0.032***
	(0.006)	(0.012)
LogMVA × DS HHI	-0.023***	-0.043***
	(0.006)	(0.011)
DS HHI	0.020**	-0.021
	(0.008)	(0.016)
LogMVA × Log DS Coverage	0.006***	0.011***
	(0.001)	(0.002)
LogMVA × IO	0.027***	0.039***
	(0.006)	(0.010)
LogMVA × IO HHI	-0.042***	-0.043***
	(0.008)	(0.016)
LogMVA × Leverage	-0.060***	-0.051***
	(0.005)	(0.010)
LogMVA × Tangibility	0.015***	-0.044***
	(0.004)	(0.006)
LogMVA × Cash	-0.102***	-0.131***
	(0.009)	(0.019)
LogMVA × Sale	0.028***	0.022***
	(0.002)	(0.004)
Log DS Coverage	0.006***	0.019***
	(0.002)	(0.004)
IO	0.003	0.020
	(0.011)	(0.021)
IO HHI	-0.044***	-0.061**
	(0.014)	(0.029)
LogAsset	0.067***	0.082***
	(0.004)	(0.009)
Leverage	-0.020**	-0.066***
	(0.009)	(0.019)
Tangibility	-0.027***	-0.038*
	(0.011)	(0.020)
Cash	0.057***	0.067**
	(0.013)	(0.028)
Sale	0.096***	0.103***
	(0.006)	(0.011)
Firm FEs	✓	✓
Time FEs	✓	✓
Obs.	46,146	34,947
Adj. R2	0.8007	0.7303

Key Findings

- **Trading Profitability**
  - ✓ Institutional investors hiring more data scientists achieve higher trading profitability
  - ✓ Each additional data scientist: 13% improvement in trading profitability
- **Strategic Portfolio Management**
  - ✓ Investors tilt portfolios toward stocks with high data scientist concentration
  - ✓ Portfolios become more concentrated to maximize data scientist advantages
- **Competitive Dynamics**
  - ✓ Investors actively track and match competitors' data scientist hiring
  - ✓ Effect stronger when hiring from key competitors
- **Market Impact**
  - ✓ High concentration of data scientists among few investors reduces price informativeness

Thank you!