Information Acquisition and Expected Returns: Evidence from EDGAR Search Traffic

Frank Weikai Li[1] Chengzhu Sun[2]

[1]Singapore Management University

[2]Hong Kong Polytechnic University

ABFER 2021

Motivation

- Information acquisition (IA)/dissemination is key to understanding asset price movements and market efficiency.
- Theoretical models of costly IA predict (Grossman and Stiglitz (1980); Verrecchia (1982); Admati (1985)):
  - IA = f((Perceived) Benefits of Info., Cost of Gathering Info.)
  - Holding the cost of IA constant, more investors will choose to become informed when they perceive greater benefits from doing so
  - Analogy: restaurants with long queue should signal delicious food (holding the location and ease of access fixed)

Motivation

- Information acquisition (IA)/dissemination is key to understanding asset price movements and market efficiency.
- Theoretical models of costly IA predict (Grossman and Stiglitz (1980); Verrecchia (1982); Admati (1985)):
  - IA = f((Perceived) Benefits of Info., Cost of Gathering Info.)
  - Holding the cost of IA constant, more investors will choose to become informed when they perceive greater benefits from doing so
  - Analogy: restaurants with long queue should signal delicious food (holding the location and ease of access fixed)
- Direct empirical evidence is sparse in financial markets
  - IA activity is difficult to observe directly (Veldkamp 2011)

Motivation (cont'd)

- Using investors' request of SEC filings through EDGAR system to measure IA :
  - EDGAR system is the main sources of firms' regulatory filings
  - Direct measure of IA for a broad cross-section of firms
  - EDGAR users tend to have higher education levels and are more likely to work in major cities with more accounting/finance jobs (Drake, Quinn, and Thornock 2017), so likely measuring sophisticated investors' IA
- **Twofold research questions**:
  - Q1:What factors drive investors' IA through EDGAR? (Determinant analysis)
  - Q2:Does costly IA reveal the value of information?

Data

- EDGAR search volume data from SEC's EDGAR log file database:
  - Information on user's unique IP address (anonymized), timestamp, searched company and type of SEC filings
  - Extensive procedure to eliminate requests made by robots or automated webcrawlers

Data

- EDGAR search volume data from SEC's EDGAR log file database:
  - Information on user's unique IP address (anonymized), timestamp, searched company and type of SEC filings
  - Extensive procedure to eliminate requests made by robots or automated webcrawlers
- We aggregate the number of unique IPs to firm-month level:
  - IP_total as the total number of unique IP addresses searching for all types of SEC filing
  - IP_funtl (IP_10K) counts only those searching for periodic accounting reports (10Q and 8K)and 10-K filing
  - Summary statistics (here)

Data

- EDGAR search volume data from SEC's EDGAR log file database:
  - Information on user's unique IP address (anonymized), timestamp, searched company and type of SEC filings
  - Extensive procedure to eliminate requests made by robots or automated webcrawlers
- We aggregate the number of unique IPs to firm-month level:
  - IP_total as the total number of unique IP addresses searching for all types of SEC filing
  - IP_funtl (IP_10K) counts only those searching for periodic accounting reports (10Q and 8K)and 10-K filing
  - Summary statistics (here)
- Monthly stock returns from CRSP, accounting data from Compustat, analyst data from I/B/E/S, institutional holdings from Thompson Reuters Financial, and securities lending data from Markit

Data Limitations

- EDGAR server log does not capture all the views/downloads that the entire universe of investors are conducting on company filings.
- Other sources of fundamental information including investor relation website, Yahoo! Finance, Bloomberg etc.

Data Limitations

- EDGAR server log does not capture all the views/downloads that the entire universe of investors are conducting on company filings.
- Other sources of fundamental information including investor relation website, Yahoo! Finance, Bloomberg etc.
- EDGAR server possesses several advantages:
  - Some firms forward investors directly to the EDGAR website to obtain their SEC filings.
  - Other sources often condense accounting information into pre-specified bins and lead to misclassification.
  - Investors could better assess firm's future prospects by reading the **qualitative** information and footnote contained in annual report (Loughran and McDonald 2011)

Time Series of Average Number of IPs

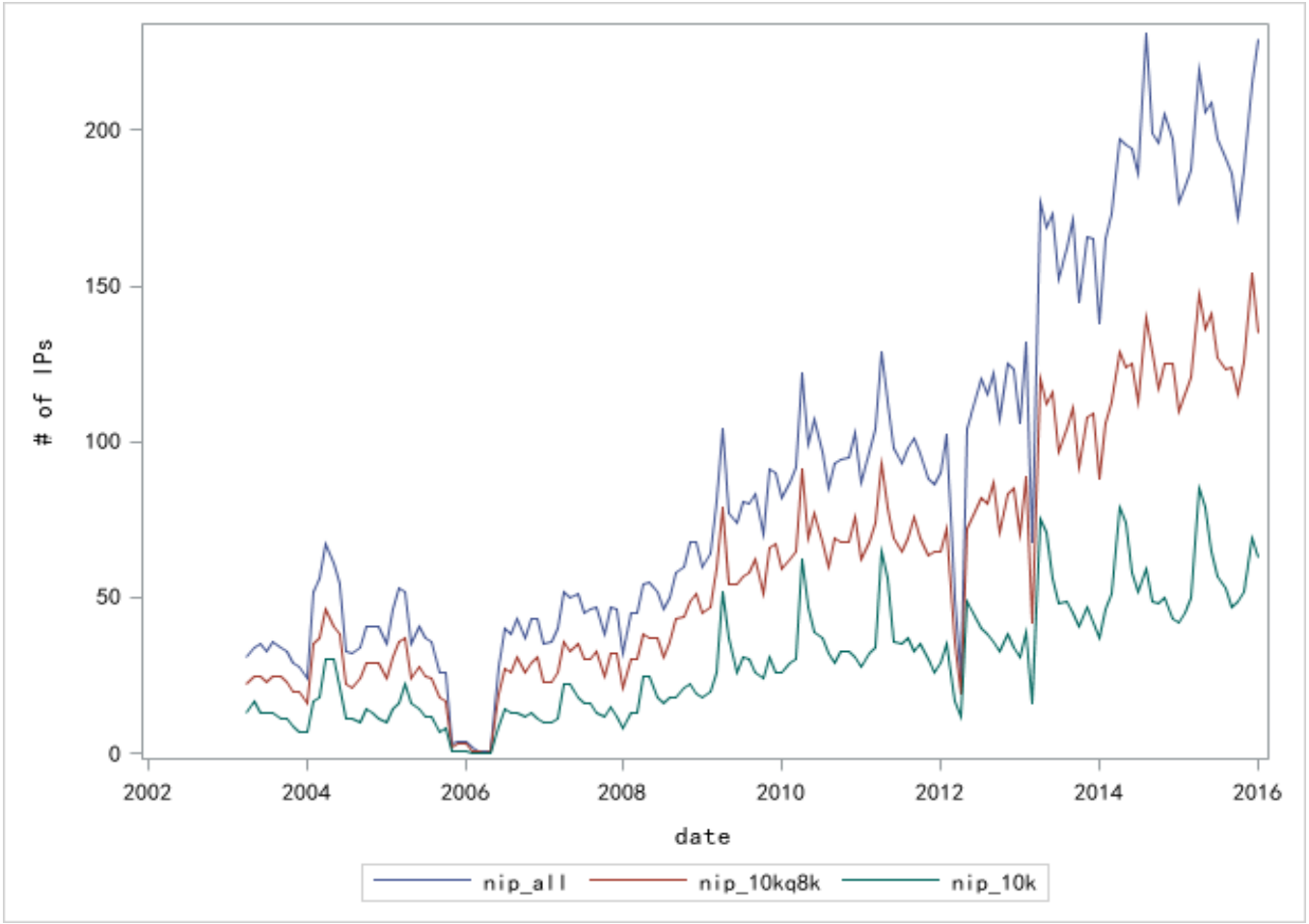
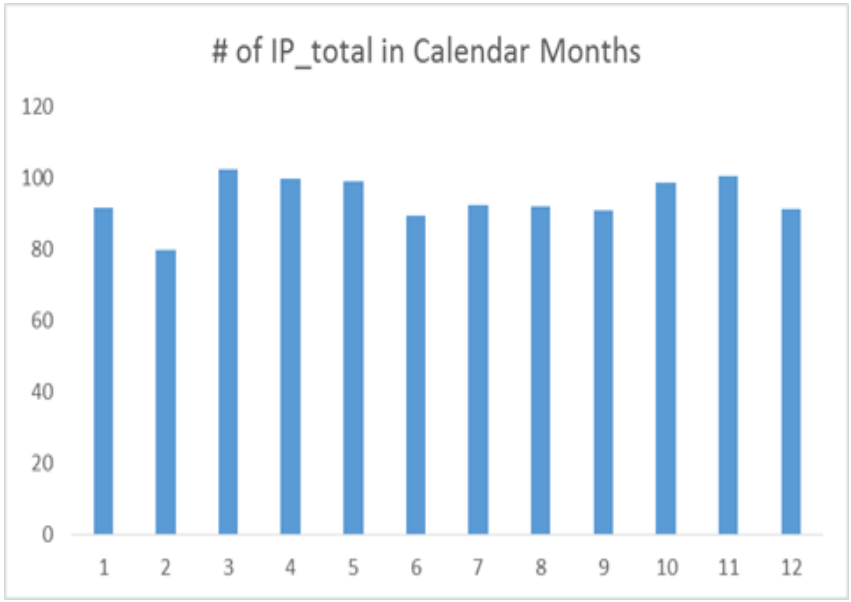
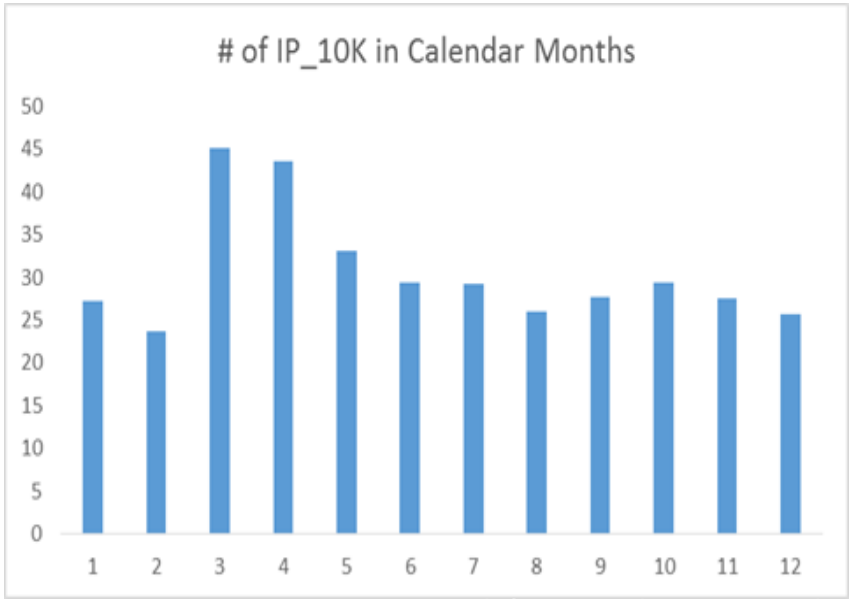


Figure: Time Series of Num of IPs for a median firm

Number of IPs in Calendar Months





Q1: What factors drive investors' IA through EDGAR?

- We develop a simple characteristics-based model of expected IA, motivated by theories:
  - Higher visibility and investor attention attract more IA as these stocks are more accessible in investors' mind
  - Information environment (positive/negative depending on whether public info complement/substitute private IA)
  - Higher valuation uncertainty attract more IA as additional information is more valuable

Q1: What factors drive investors' IA through EDGAR?

- We develop a simple characteristics-based model of expected IA, motivated by theories:
  - Higher visibility and investor attention attract more IA as these stocks are more accessible in investors' mind
  - Information environment (positive/negative depending on whether public info complement/substitute private IA)
  - Higher valuation uncertainty attract more IA as additional information is more valuable
- Following the literature, we use:
  - Size proxy for investor visibility
  - Trading volume proxy for investor attention (Gervais, Kaniel, and Mingelgrin (2001); Barber and Odean (2008))
  - Analyst coverage proxy for information environment (Hong, Lim, and Stein (2000))
  - Idiosyncratic volatility proxy for valuation uncertainty (Zhang (2006))

Cross-Sectional Determinants of # of IPs-Results

- $Log(1 + IP_{i,t}) = \beta_0 + \beta_1 LnME_{i,t} + \beta_2 Coverage_{i,t} + \beta_3 Turnover12_{i,t} + \beta_4 IVOL_{i,t} + \epsilon_{i,t}$

Panel A: Dependent Variable is log(1+# of unique IP adresses searching all EDGAR files)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LnME	0.2713***	0.2356***	0.2475***	0.2943***	0.2992***	0.3015***	0.3026***	0.2608***	0.2628***
	(69.44)	(71.54)	(73.46)	(75.60)	(76.94)	(77.29)	(77.58)	(75.05)	(74.98)
Coverage		0.1310***	0.0422***	0.0382***	0.0321***	0.0332***	0.0360***	0.0337***	0.0399***
		(32.65)	(14.39)	(14.36)	(12.17)	(12.56)	(14.17)	(13.99)	(16.86)
Turnover12			1.0083***	0.7934***	0.7862***	0.7912***	0.7877***	0.8175***	0.8113***
			(30.21)	(29.08)	(30.04)	(29.75)	(30.52)	(30.68)	(30.92)
Ivol				9.1266***	9.0159***	9.0510***	9.0215***	8.5748***	8.0871***
				(34.65)	(33.38)	(33.16)	(32.36)	(31.55)	(31.65)
Mom					-0.0518***	-0.0529***	-0.0507***	-0.0508***	-0.0513***
					(-6.00)	(-6.19)	(-5.99)	(-6.38)	(-6.43)
LnBM						0.0171***	0.0158***	0.0087***	0.0108***
						(8.19)	(7.25)	(4.06)	(5.16)
IO							-0.0299**	0.0657***	0.0575***
							(-1.99)	(4.80)	(4.37)
SP500								0.3634***	0.3591***
								(58.81)	(58.60)
EAM									0.1587***
									(9.62)
Constant	2.5352***	2.6342***	2.5357***	2.0730***	2.0483***	2.0408***	2.0449***	2.2164***	2.1892***
	(39.20)	(40.68)	(40.19)	(33.45)	(33.37)	(33.32)	(32.62)	(34.26)	(34.03)
Ave.R-sq	0.404	0.483	0.520	0.554	0.558	0.559	0.563	0.574	0.582
N.of Obs.	610651	488129	488129	488123	488123	488123	484835	484835	484835

Q2: Abnormal IA and Future Stock Returns

- Q2: Does costly information acquisition reveal investors' expected payoffs from using the information?
  - In theory, the direction of the information could be either positive or negative
  - In reality due to short-sale constraints, investors more likely engage in costly IA when expected payoff is positive

Q2: Abnormal IA and Future Stock Returns

- Q2: Does costly information acquisition reveal investors' expected payoffs from using the information?
  - In theory, the direction of the information could be either positive or negative
  - In reality due to short-sale constraints, investors more likely engage in costly IA when expected payoff is positive
- To test:
  - Abnormal number of IPs (AIP) proxy for abnormal IA
  - At month end, we sort stocks into deciles based on AIP and compute the Carhart(1997) four-factor alphas of each decile portfolio over the next month

EW portfolio strategies based on AIP

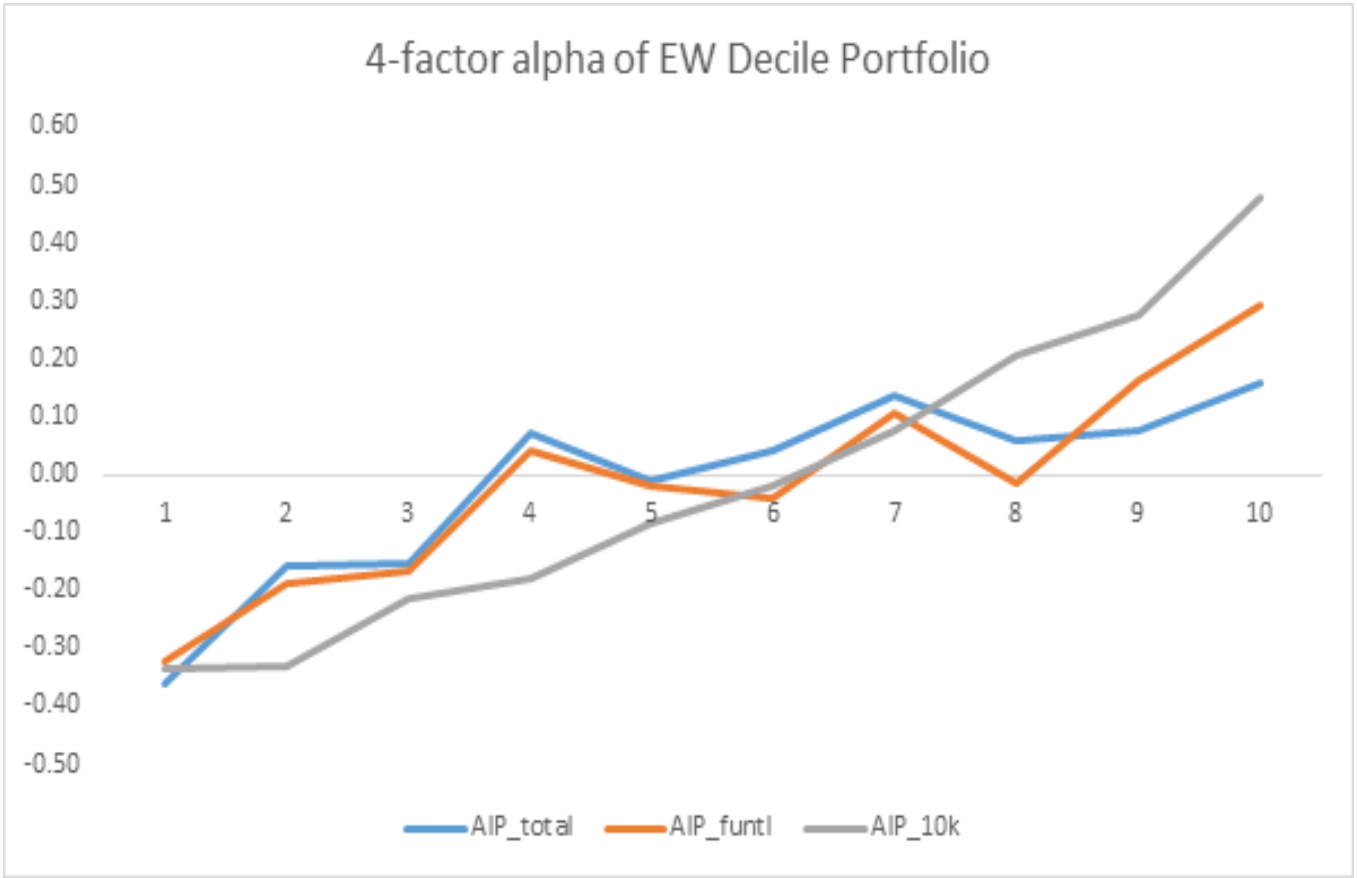


Figure: 4-factor alpha of EW Decile Portfolios, 2003-2014

- L/S portfolio based on AIP_10K generates 4-factor alpha of 82 basis points (t-stat=4.35).

VW portfolio strategies based on AIP

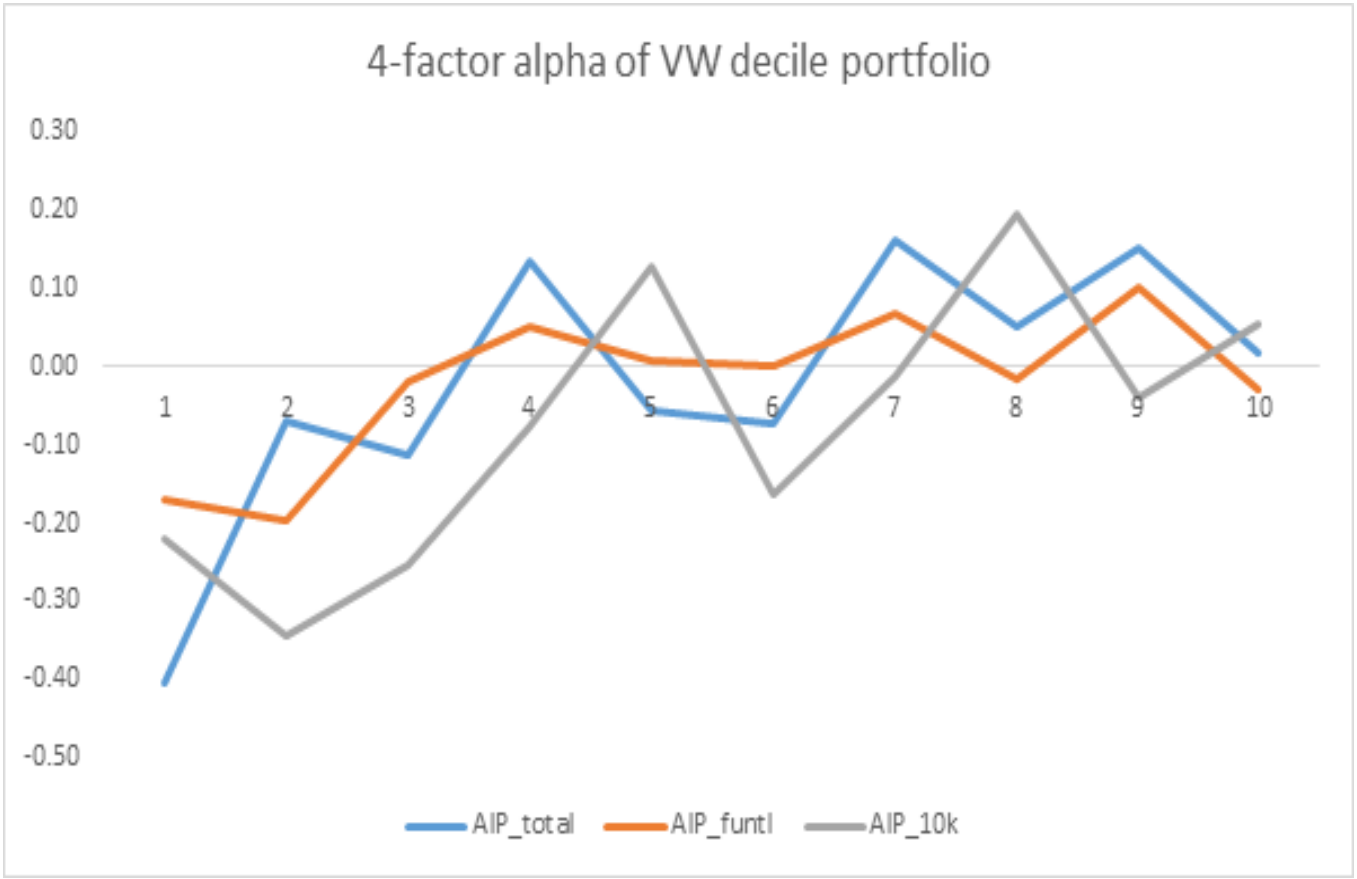
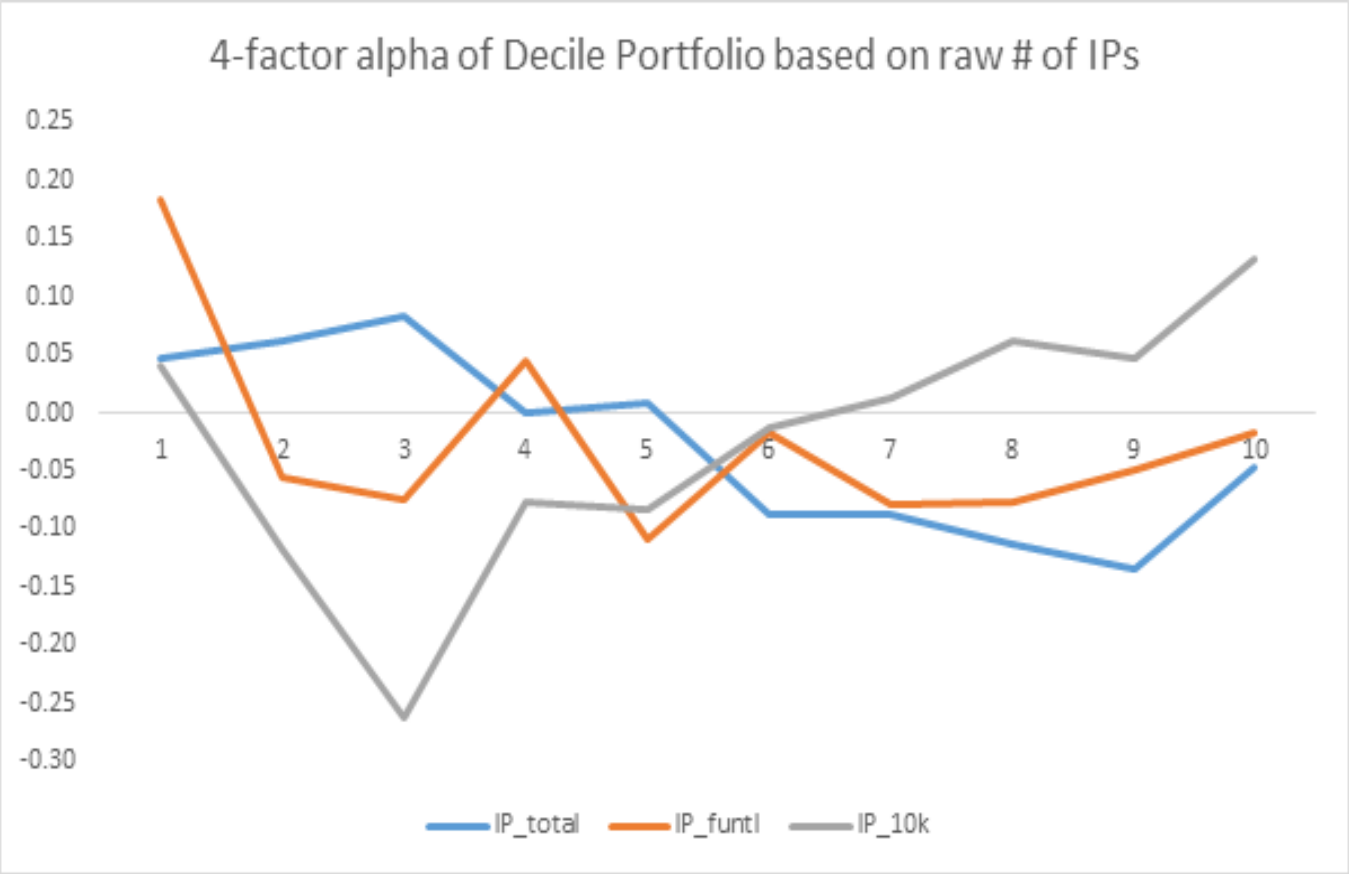


Figure: 4-factor alpha of VW Decile Portfolios, 2003-2014

Returns to Portfolios Sorted on Raw Number of IPs

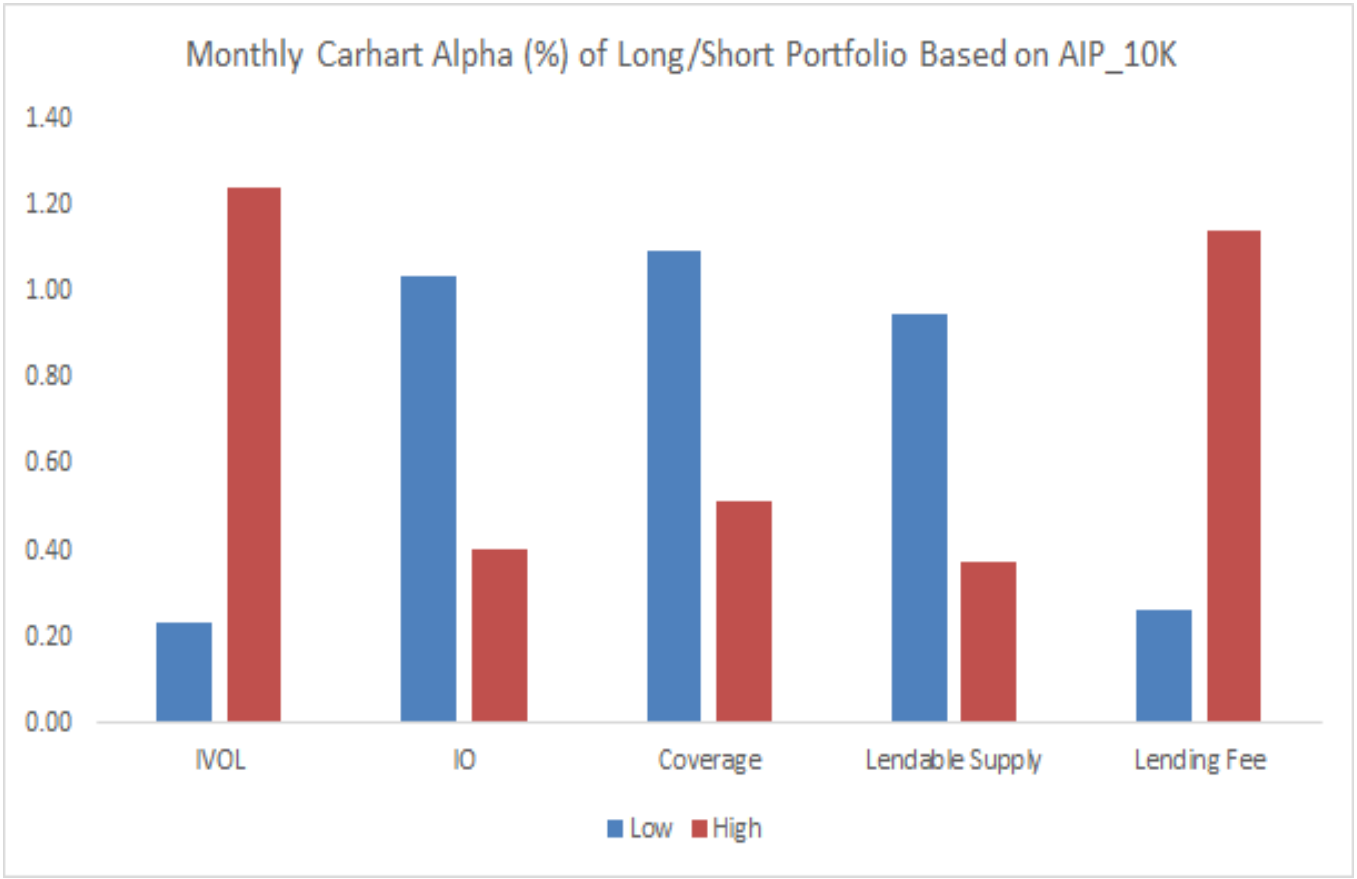


- Important to control for the expected level of information acquisition when uncovering investors' expected payoff.
- Large raw number of IPs could be driven by low costs of information acquisition, rather than high benefits (restaurants with long queue due to good location not necessarily delicious food).

Robustness of Portfolio Sorts

	EW	VW
Gross return-weighed portfolio	1.096	NA
	(5.16)	
DGTW adjusted	0.910	0.410
	(4.51)	(2.22)
FF48 Industry-adjusted	0.739	0.155
	(3.26)	(1.16)
FF + Cahart + PS Factor	0.800	0.348
	(4.23)	(1.78)
FF five factors (2015)	0.685	0.248
	(3.36)	(1.19)
Mispricing factors (Stambaugh and Yuan 2017)	0.892	0.276
	(4.42)	(1.35)
Q-factor (Hou, Xue and Zhang 2015)	0.897	0.183
	(4.66)	(0.87)
Remove microcap stocks	0.518	0.276
	(2.58)	(1.35)
Skip six months	0.532	0.266
	(2.23)	(1.28)
2003-2008	0.620	0.261
	(2.41)	(0.89)
2009-2014	1.073	0.121
	(3.74)	(0.45)

Short-Sale Constraints and Arbitrage Frictions



- Positive expected return information in AIP is more pronounced among stocks with higher short-sale constraints and arbitrage frictions.

Complexity of SEC Filings

Panel A: Double sort on residual file size and AIP_10K

	Low AIP_10K	High AIP_10K	High-Low
Large Filing Size	-0.65	0.59	1.24
	(-4.22)	(3.23)	(4.88)
Small Filing Size	-0.27	0.38	0.66
	(-1.84)	(2.68)	(3.30)
Large Filing Size - Small Filing Size			0.58
			(2.29)

Panel B: Double sort on word count and AIP_10K

	Low AIP_10K	High AIP_10K	High-Low
More word count	-0.48	0.52	1.00
	(-3.29)	(2.39)	(5.06)
Lesser word count	-0.36	0.20	0.56
	(-3.02)	(1.35)	(2.93)
More word count - Lesser word count			0.44
			(1.99)

- Use residual file size and word counts in 10-K to proxy for complexity (You and Zhang 2009; Loughran and McDonald 2014)
- Analogy: restaurants with long queue despite located far away should indicate very delicious food

Cross-sectional Variation at IP level

Panel A: EDGAR searching for current and historical filings

	Low AIP_10K	High AIP_10K	High-Low
Current filings	-0.41	0.21	0.61
	(-2.29)	(1.29)	(3.08)
Both current and historical filings	-0.45	0.55	1.00
	(-4.54)	(3.63)	(5.28)
Both current and historical filings - Current filings			0.39
			(2.53)

Panel B: Daytime and Nighttime searches

	Low AIP_10K	High AIP_10K	High-Low
Nighttime search	-0.39	0.43	0.82
	(-3.25)	(2.92)	(4.71)
Daytime search	-0.35	0.45	0.79
	(-3.23)	(3.15)	(4.70)
Nighttime search - Daytime search			0.03
			(0.25)

Fama-MacBeth Regressions

	All EDGAR Filings			10-K, 10-Q and 8-K			10-K		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
AIP	0.0060***	0.0053***	0.0050***	0.0047***	0.0041***	0.0042***	0.0051***	0.0046***	0.0044***
	(2.68)	(2.64)	(2.88)	(2.70)	(2.78)	(2.94)	(3.73)	(3.81)	(3.74)
Rev		-0.0247***	-0.0283***		-0.0245***	-0.0281***		-0.0247***	-0.0284***
		(-3.18)	(-3.74)		(-3.16)	(-3.72)		(-3.19)	(-3.75)
LnME		-0.0006	-0.0014**		-0.0006	-0.0014**		-0.0006	-0.0014**
		(-0.89)	(-2.59)		(-0.92)	(-2.60)		(-0.93)	(-2.58)
LnBM		0.0019	0.0014		0.0019	0.0013		0.0019	0.0013
		(1.64)	(1.29)		(1.59)	(1.24)		(1.58)	(1.24)
Mom		-0.0058	-0.0048		-0.0057	-0.0047		-0.0058	-0.0048
		(-0.95)	(-0.88)		(-0.94)	(-0.86)		(-0.94)	(-0.86)
Ivol			-0.0015			-0.0025			-0.0007
			(-0.02)			(-0.04)			(-0.01)
Turnover12			-0.0094			-0.0091			-0.0089
			(-1.37)			(-1.32)			(-1.28)
IO			0.0122***			0.0119***			0.0114***
			(4.00)			(3.94)			(3.86)
Constant	0.0123**	0.0122	0.0119**	0.0122**	0.0122*	0.0120**	0.0122**	0.0123*	0.0119**
	(2.18)	(1.65)	(2.33)	(2.18)	(1.66)	(2.36)	(2.18)	(1.67)	(2.35)
Ave.R-sq	0.003	0.030	0.046	0.003	0.030	0.046	0.003	0.030	0.046
N.of Obs.	483667	483667	480793	483667	483667	480793	483667	483667	480793

- $Ret_{i,t+1} = \beta_0 + \beta_1 AIP_{i,t} + \gamma X_{i,t} + \epsilon_{i,t}$
- Coefficient of AIP_10K implies a monthly return spread of 105 basis points between extreme deciles

Channels of Return Predictability

- Under short-sale constraints, investors rationally allocate more effort toward undervalued firms with large price appreciation potential.

Channels of Return Predictability

- Under short-sale constraints, investors rationally allocate more effort toward undervalued firms with large price appreciation potential.
- Two non-mutually exclusive channels to identify mispricing:
  - Investors identify firms with improving fundamentals
  - Investors observe shocks to stock prices that are unwarranted by firms' fundamentals

Channel 1: Predicting Fundamental Performance

- Change of quarterly Return-on-Assets (ROA) from four quarters ago
- Standardized unexpected earnings (SUE)
- Monthly forecast revision of analysts' consensus annual EPS forecast
- Results help rule out change of visibility (Merton 1987) or risk-based explanations

	Change of ROA			SUE			Forecast Revision		
	AIP_total	AIP_fundl	AIP_10K	AIP_total	AIP_fundl	AIP_10K	AIP_total	AIP_fundl	AIP_10K
AIP	0.0017*	0.0026**	0.0028***	0.0013	0.0026**	0.0043***	0.0007***	0.0016***	0.0019***
	(1.96)	(2.51)	(2.92)	(1.42)	(2.22)	(3.57)	(2.78)	(6.19)	(5.28)
LROA	-0.3425***	-0.3428***	-0.3430***						
	(-4.71)	(-4.73)	(-4.74)						
LnME	0.0008	0.0008	0.0008	-0.0022***	-0.0021***	-0.0021***	-0.0005	-0.0005	-0.0005
	(1.27)	(1.31)	(1.33)	(-3.73)	(-3.68)	(-3.63)	(-1.51)	(-1.55)	(-1.63)
LnBM	-0.0013	-0.0012	-0.0012	-0.0009	-0.0009	-0.0008	-0.0008**	-0.0008**	-0.0009**
	(-0.87)	(-0.84)	(-0.83)	(-0.50)	(-0.48)	(-0.44)	(-2.37)	(-2.39)	(-2.47)
Mom	0.0100***	0.0099***	0.0100***	0.0220***	0.0220***	0.0219***	0.0025***	0.0025***	0.0025***
	(3.55)	(3.56)	(3.57)	(8.19)	(8.17)	(8.18)	(5.20)	(5.14)	(5.18)
Cov	0.0004	0.0004	0.0005	0.0002	0.0003	0.0003	0.0021***	0.0021***	0.0021***
	(0.29)	(0.31)	(0.32)	(0.23)	(0.27)	(0.29)	(3.30)	(3.29)	(3.28)
Turnover12	-0.0118**	-0.0117**	-0.0117**	0.0316***	0.0318***	0.0319***	-0.0082***	-0.0082***	-0.0082***
	(-2.43)	(-2.42)	(-2.43)	(3.45)	(3.47)	(3.47)	(-3.15)	(-3.16)	(-3.17)
IO	-0.0010	-0.0011	-0.0013	-0.0093***	-0.0095***	-0.0096***	0.0049***	0.0051***	0.0052***
	(-0.48)	(-0.51)	(-0.61)	(-3.75)	(-3.87)	(-3.97)	(5.47)	(5.61)	(5.73)
IVOL	-0.0777	-0.0773	-0.0775	0.2287**	0.2330**	0.2365**	-0.1111**	-0.1115**	-0.1138**
	(-1.41)	(-1.41)	(-1.42)	(2.29)	(2.32)	(2.34)	(-2.35)	(-2.37)	(-2.41)
Time FE	yes	yes	yes	yes	yes	yes	yes	yes	yes
Adj.R-sq	0.056	0.056	0.056	0.023	0.023	0.023	0.002	0.002	0.002
N.of Obs.	128504	128504	128504	150712	150712	150712	348130	348130	348130

Channel 2:Mutual Fund Outflows Induced Underpricing

- Mutual fund outflow-induced fire-sale as exogeneous shock to stock price
- Outflow is calculated following Edmans, Goldstein, and Jiang (2012)

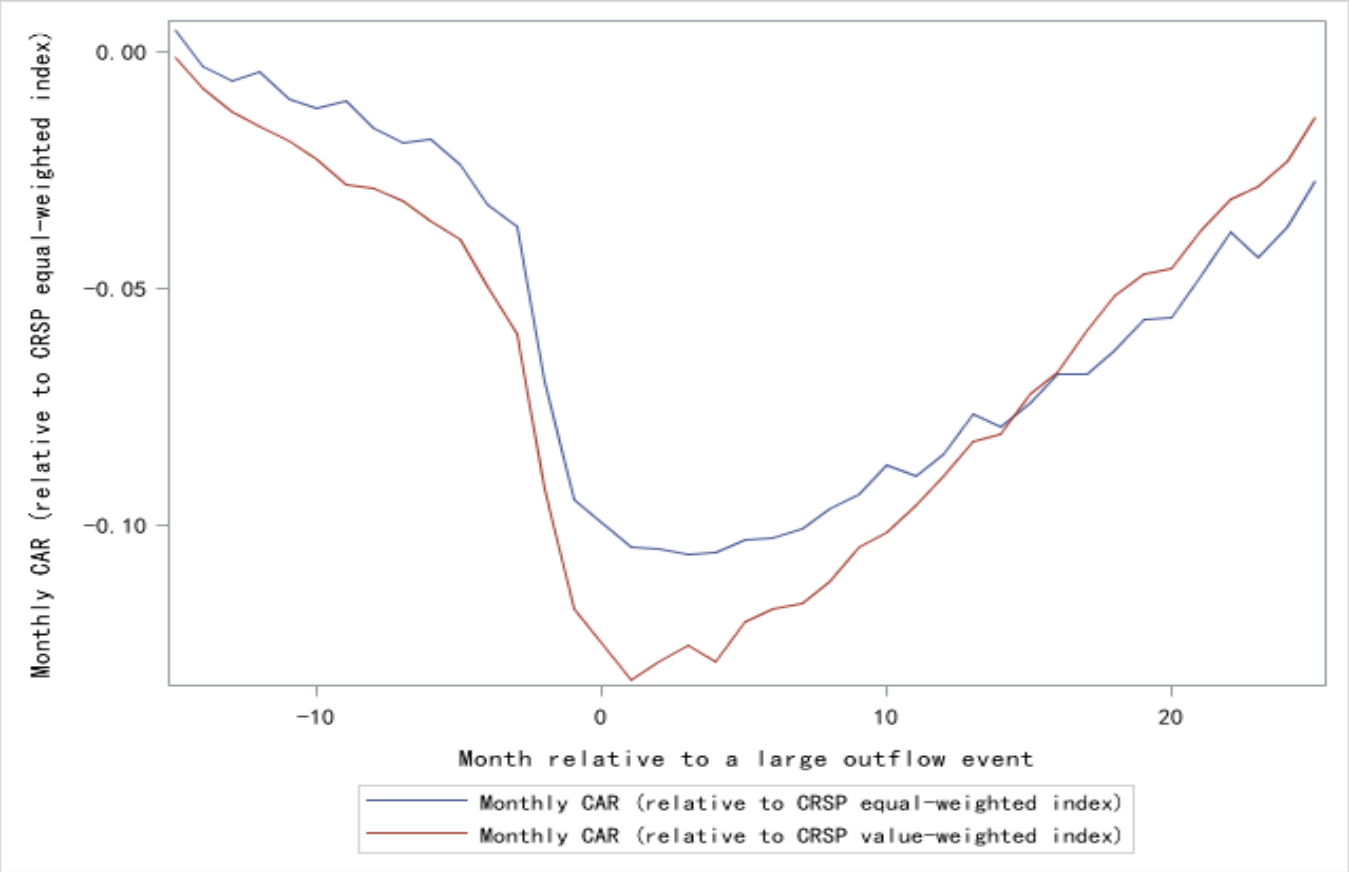


Figure: Effect of Mutual Funds Hypothetical Sales on Stock Prices

Channel 2: Mutual Fund Outflows Induced Underpricing

- $dAIP_{i,q+1} = \beta_0 + \beta_1 Outflows_{i,q} + \beta_2 X_{i,q} + \epsilon_{i,q+1}$

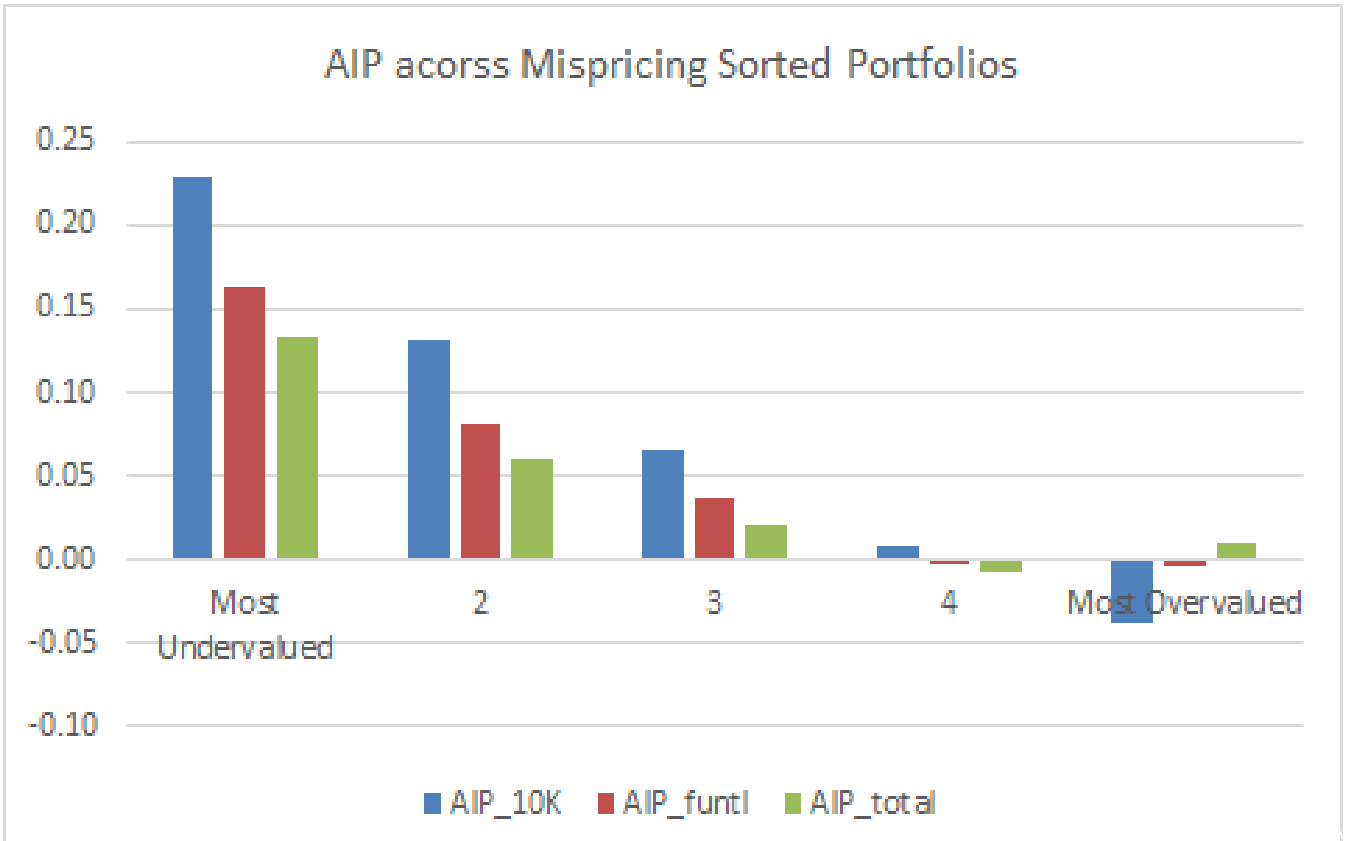
	dAIP_total		dAIP_funtl		dAIP_10K	
	(1)	(2)	(3)	(4)	(5)	(6)
Outflows	-2.4242***	-1.7256***	-1.9145***	-1.3527***	-1.9303**	-1.5459**
	(-4.02)	(-4.92)	(-3.36)	(-3.27)	(-2.06)	(-2.31)
LnME		-0.0091***		-0.0094***		-0.0093***
		(-6.03)		(-5.68)		(-5.81)
LnBM		0.0013		-0.0014		-0.0017
		(0.56)		(-0.57)		(-0.75)
Coverage		0.0080***		0.0076***		0.0087***
		(4.50)		(4.28)		(3.70)
Ivol		-1.8233***		-1.9963***		-1.8354***
		(-6.48)		(-7.68)		(-6.19)
Turnover12		-0.0015		0.0158		0.0203
		(-0.09)		(1.13)		(1.56)
IO		-0.0023		-0.0141**		-0.0143**
		(-0.36)		(-2.54)		(-2.28)
Mom		-0.0336***		-0.0370***		-0.0398***
		(-5.17)		(-5.70)		(-7.68)
Constant	0.0007	0.0901***	0.0050**	0.1036***	0.0049**	0.0967***
	(0.29)	(7.79)	(2.09)	(8.54)	(2.06)	(6.54)
Ave.R-sq	0.001	0.031	0.001	0.034	0.001	0.026
N.of Obs.	131863	131041	131863	131041	131863	131041

EDGAR Search and Mispricing

- If investors already know which firms are potentially undervalued even before analyzing SEC filings, what is the incremental value of acquiring information through EDGAR?
- Conjecture: acquiring fundamental information through EDGAR could help investors identify truely mispriced stocks.
- Use the composite mispricing measure constructed by Stambaugh, Yu, and Yuan (2015) to identify mispricing

EDGAR Search and Mispricing

- If investors already know which firms are potentially undervalued even before analyzing SEC filings, what is the incremental value of acquiring information through EDGAR?
- Conjecture: acquiring fundamental information through EDGAR could help investors identify truely mispriced stocks.
- Use the composite mispricing measure constructed by Stambaugh, Yu, and Yuan (2015) to identify mispricing



EDGAR Search and Mispricing (2)

- Accessing EDGAR filings could help investors identify truely mispriced stocks among those with similar mispricing measures.

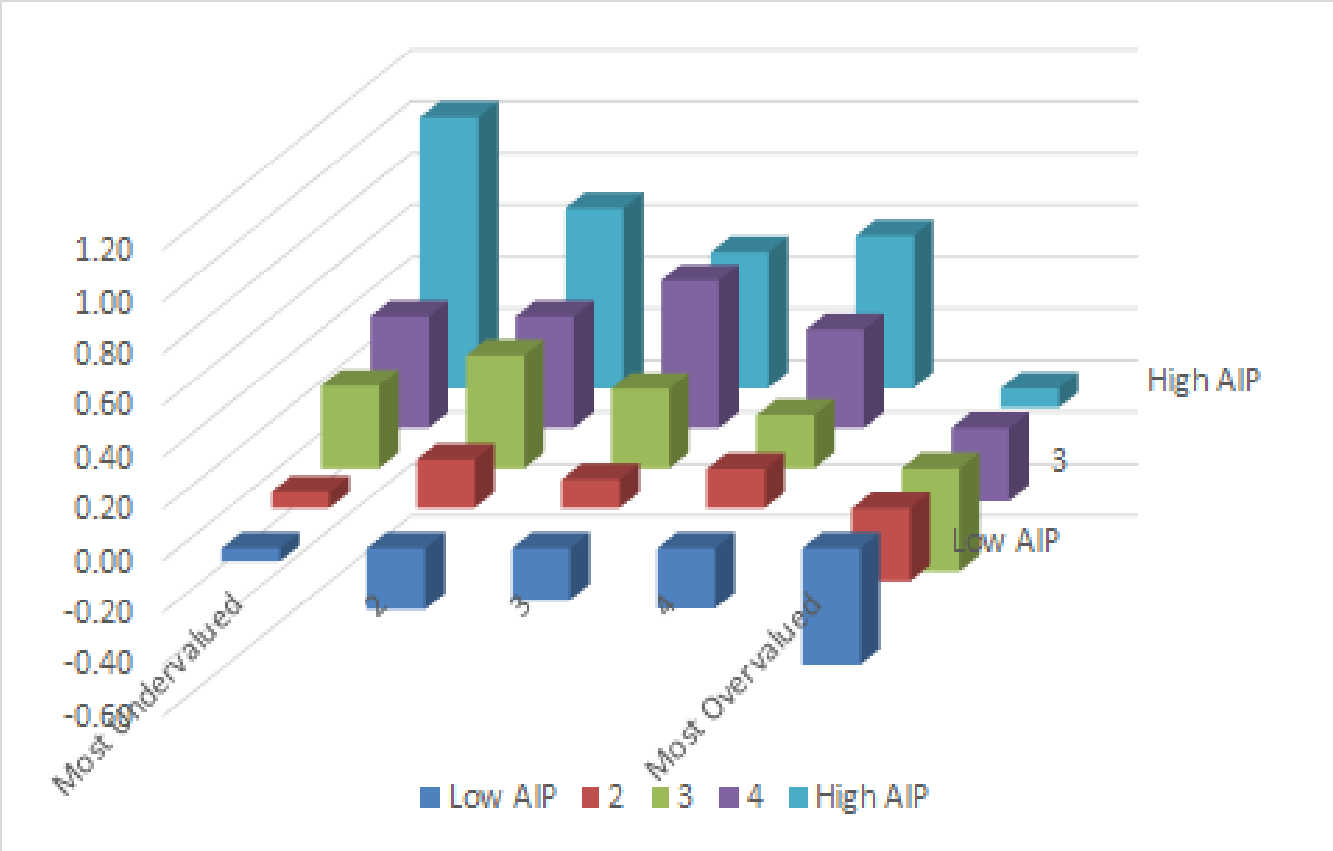


Figure: Two-way Independent Sorts on AIP and Composite Mispricing Measure

Ruling out alternative explanations

- Firm Events(here)
- Breadth of Ownership and Extreme Returns (here)
- News Coverage(here)
- Attention-Driven Price Pressure (here)
- Investor Recognition
- Omitted risk factors (here)

EDGAR Search and Investor Trading

	Net Purchases by Mutual Funds			Retail Order Imbalance		
	(1)	(2)	(3)	(4)	(5)	(6)
AIP_total	0.0030			0.0090***		
	(0.52)			(7.16)		
AIP_funtl		0.0046			0.0079***	
		(0.77)			(6.89)	
AIP_10K			0.0065			0.0076***
			(1.00)			(7.81)
LnME	-0.0003	-0.0003	-0.0003	-0.0002	-0.0002	-0.0002
	(-1.07)	(-1.08)	(-1.06)	(-0.20)	(-0.22)	(-0.25)
LnBM	0.0003	0.0004	0.0003	0.0045***	0.0045***	0.0044***
	(0.32)	(0.38)	(0.34)	(4.95)	(4.96)	(4.84)
Coverage	-0.0008	-0.0007	-0.0005	-0.0025***	-0.0026***	-0.0027***
	(-0.30)	(-0.26)	(-0.22)	(-3.46)	(-3.61)	(-3.76)
IVOL	-0.1965*	-0.1975*	-0.2012*	-0.3268***	-0.3257***	-0.3243***
	(-1.96)	(-1.94)	(-1.95)	(-6.41)	(-6.40)	(-6.37)
Turnover12	-0.0069**	-0.0068**	-0.0064**	-0.0350***	-0.0351***	-0.0352***
	(-2.04)	(-2.22)	(-2.38)	(-12.83)	(-12.79)	(-12.75)
IO	0.0739***	0.0736***	0.0728***	0.0109***	0.0117***	0.0124***
	(2.86)	(2.90)	(2.92)	(3.01)	(3.26)	(3.42)
MOM	0.0066***	0.0067***	0.0065***	-0.0030	-0.0031*	-0.0031*
	(6.44)	(6.29)	(7.74)	(-1.67)	(-1.73)	(-1.70)
Constant	0.0048*	0.0050*	0.0053*	0.0469***	0.0467***	0.0465***
	(1.95)	(1.91)	(1.96)	(6.63)	(6.61)	(6.58)
Ave.R-sq	0.113	0.113	0.113	0.010	0.010	0.010
N.of Obs.	131795	131795	131795	184715	184715	184715

- More information acquisition activities through the EDGAR system leads to significant net buying from retail investor, but not mutual funds
- EDGAR searches mainly capture individual investors' information acquisition

Conclusion

- Use a novel dataset containing investors' access of EDGAR filings to measure IA.
- Abnormal number of IPs searching for EDGAR filings strongly predict future returns and fundamentals
- Provide direct evidence supporting theoretical models of endogenous information acquistion.
- Highlights the promise of using the collective wisdom of investors to study important economic outcomes.

Firm events, Breadth of Ownership and Extreme Returns

	AIP_total	AIP_fundl	AIP_10K	AIP_total	AIP_fundl	AIP_10K	AIP_total	AIP_fundl	AIP_10K
AIP	0.0041**	0.0045***	0.0043***	0.0042**	0.0047***	0.0043***	0.0053***	0.0047***	0.0046***
	(2.45)	(3.09)	(3.81)	(2.49)	(3.10)	(3.94)	(3.38)	(3.14)	(4.20)
Rev	-0.0312***	-0.0309***	-0.0312***	-0.0316***	-0.0312***	-0.0315***	-0.0352***	-0.0351***	-0.0358***
	(-4.26)	(-4.23)	(-4.27)	(-4.34)	(-4.29)	(-4.34)	(-4.46)	(-4.48)	(-4.54)
LnME	-0.0018***	-0.0018***	-0.0018***	-0.0018***	-0.0018***	-0.0018***	-0.0018***	-0.0018***	-0.0017***
	(-3.69)	(-3.74)	(-3.72)	(-3.69)	(-3.72)	(-3.72)	(-3.67)	(-3.71)	(-3.73)
LnBM	0.0016	0.0015	0.0015	0.0016	0.0015	0.0015	0.0015	0.0014	0.0014
	(1.50)	(1.44)	(1.46)	(1.52)	(1.46)	(1.47)	(1.48)	(1.43)	(1.41)
Mom	-0.0065	-0.0064	-0.0064	-0.0065	-0.0064	-0.0063	-0.0065	-0.0065	-0.0064
	(-1.15)	(-1.14)	(-1.12)	(-1.14)	(-1.12)	(-1.11)	(-1.16)	(-1.16)	(-1.14)
Ivol	0.0169	0.0131	0.0174	0.0240	0.0209	0.0220	-0.0636	-0.0692	-0.0768
	(0.24)	(0.18)	(0.24)	(0.34)	(0.29)	(0.31)	(-0.69)	(-0.74)	(-0.78)
Turnover12	-0.0087	-0.0082	-0.0084	-0.0091	-0.0086	-0.0089	-0.0085	-0.0079	-0.0080
	(-1.25)	(-1.18)	(-1.19)	(-1.29)	(-1.22)	(-1.24)	(-1.22)	(-1.13)	(-1.14)
IO	0.0118***	0.0113***	0.0110***	0.0120***	0.0115***	0.0112***	0.0120***	0.0114***	0.0111***
	(3.58)	(3.52)	(3.40)	(3.55)	(3.50)	(3.37)	(3.56)	(3.51)	(3.38)
SUE	0.0028***	0.0028***	0.0027***	0.0028***	0.0028***	0.0027***	0.0027***	0.0028***	0.0027***
	(8.48)	(8.52)	(8.57)	(8.57)	(8.62)	(8.64)	(8.49)	(8.53)	(8.54)
EAM	0.0033***	0.0035***	0.0028**	0.0031**	0.0033**	0.0028**	0.0031**	0.0032**	0.0027**
	(2.61)	(2.69)	(2.33)	(2.55)	(2.60)	(2.31)	(2.51)	(2.56)	(2.27)
Upgrade	0.0023***	0.0023***	0.0025***	0.0023***	0.0023***	0.0024***	0.0024***	0.0024***	0.0025***
	(2.76)	(2.76)	(2.95)	(2.79)	(2.77)	(2.94)	(2.89)	(2.90)	(3.03)
Downgrade	-0.0010	-0.0011	-0.0013	-0.0009	-0.0010	-0.0012	-0.0013	-0.0012	-0.0015*
	(-1.00)	(-1.16)	(-1.38)	(-0.90)	(-1.03)	(-1.29)	(-1.54)	(-1.36)	(-1.78)
DM	0.0030***	0.0031***	0.0031***	0.0031***	0.0032***	0.0031***	0.0031***	0.0031***	0.0031***
	(2.78)	(2.77)	(2.75)	(2.95)	(2.96)	(2.86)	(2.87)	(2.89)	(2.83)
num_8K				-0.0010	-0.0012*	-0.0004	-0.0010	-0.0012*	-0.0004
				(-1.55)	(-1.80)	(-0.64)	(-1.51)	(-1.76)	(-0.63)
dBreadth							0.0722	0.0825	0.0836
							(0.94)	(1.06)	(1.11)
Maxret							-0.0308	-0.0317	-0.0346
							(-1.52)	(-1.60)	(-1.53)
Constant	0.0121**	0.0124**	0.0123**	0.0125**	0.0128**	0.0125**	0.0123**	0.0127**	0.0124**
	(2.46)	(2.52)	(2.50)	(2.50)	(2.56)	(2.53)	(2.41)	(2.48)	(2.46)
Ave.R-sq	0.053	0.053	0.053	0.054	0.054	0.053	0.057	0.057	0.057
N.of Obs.	443261	443261	443261	443261	443261	443261	442698	442698	442698

Predicting Long-horizon Returns

	Ret(2,4)	Ret(5,7)	Ret(8,13)	Ret(14,25)
AIP_10K	0.0102***	0.0068**	0.0150	0.0175
	(2.95)	(2.05)	(1.57)	(0.64)
Rev	-0.0072	0.0037	0.0033	-0.0451
	(-0.53)	(0.21)	(0.11)	(-0.93)
LnME	-0.0023	-0.0013	-0.0015	-0.0048
	(-1.64)	(-1.03)	(-0.61)	(-1.11)
LnBM	0.0046*	0.0041	0.0118**	0.0197*
	(1.72)	(1.57)	(2.36)	(1.79)
Mom	-0.0193	-0.0117	-0.0300*	-0.0421
	(-1.24)	(-0.88)	(-1.75)	(-1.26)
Ivol	0.0407	-0.0184	0.2652	0.5759
	(0.20)	(-0.10)	(0.73)	(0.84)
Turnover12	-0.0165	-0.0312*	-0.0451	-0.0488
	(-0.92)	(-1.95)	(-1.53)	(-1.08)
IO	0.0116	0.0152**	0.0414**	0.0956**
	(1.63)	(2.18)	(2.42)	(2.47)
Constant	0.0370**	0.0281*	0.0451	0.0947
	(2.41)	(1.72)	(1.53)	(1.51)
Ave.R-sq	0.051	0.044	0.036	0.035
N.of Obs.	469185	456068	425505	360584

- Return predictability of AIP not from transitory price pressure that reverses subsequently (Da, Engelberg and Gao, 2011) (back to altern)

Controlling for News Coverage and News Sentiment

	(1)	(2)	(3)	(4)	(5)	(6)
AIP_total	0.0043**			0.0041**		
	(2.33)			(2.40)		
AIP_funtl		0.0040**			0.0044***	
		(2.13)			(2.86)	
AIP_10K			0.0050***			0.0052***
			(3.65)			(3.95)
REV	-0.0250***	-0.0232***	-0.0252***	-0.0270***	-0.0267***	-0.0276***
	(-2.95)	(-2.83)	(-2.96)	(-3.21)	(-3.18)	(-3.24)
LnME	-0.0013***	-0.0013**	-0.0012**	-0.0015***	-0.0016***	-0.0015***
	(-2.62)	(-2.60)	(-2.27)	(-2.67)	(-2.74)	(-2.72)
LnBM	0.0011	0.0012	0.0012	0.0006	0.0005	0.0003
	(1.03)	(1.05)	(1.05)	(0.59)	(0.45)	(0.27)
MOM	-0.0053	-0.0051	-0.0052	-0.0050	-0.0049	-0.0046
	(-0.84)	(-0.81)	(-0.81)	(-0.75)	(-0.74)	(-0.69)
IVOL	0.1334*	0.1171	0.1363*	0.1338*	0.1290*	0.1396*
	(1.66)	(1.52)	(1.68)	(1.73)	(1.69)	(1.74)
Turnover12	-0.0083	-0.0093	-0.0070	-0.0116	-0.0110	-0.0105
	(-0.96)	(-1.01)	(-0.82)	(-1.11)	(-1.08)	(-1.05)
IO	0.0082**	0.0084***	0.0070**	0.0110***	0.0111***	0.0105***
	(2.56)	(2.82)	(2.12)	(3.74)	(3.71)	(3.61)
News Coverage	-0.0004	-0.0005	-0.0005			
	(-0.66)	(-0.71)	(-0.72)			
News Sentiment				0.0161***	0.0161***	0.0171***
				(3.88)	(3.86)	(3.48)
Constant	0.0153***	0.0155***	0.0153***	0.0122**	0.0122**	0.0115**
	(2.88)	(2.92)	(2.91)	(2.26)	(2.27)	(2.10)
Ave.R-sq	0.055	0.055	0.055	0.056	0.055	0.056
N.of Obs.	264816	264816	264816	264816	264816	264816

- Control for news coverage and news sentiment using RavenPack (back to altern)

Predicting Earnings Annoucement Returns

	Market-adjusted CAR(-1,+1)			DGTW-adjusted CAR(-1,+1)		
	AIP_total	AIP_fundl	AIP_10K	AIP_total	AIP_fundl	AIP_10K
AIP	0.0020	0.0025*	0.0036***	0.0019	0.0024*	0.0033***
	(1.39)	(1.90)	(2.74)	(1.45)	(1.93)	(2.93)
Rev	-0.0001	0.0003	0.0002	0.0000	0.0004	0.0004
	(-0.02)	(0.13)	(0.08)	(0.01)	(0.17)	(0.19)
LnME	0.0001	0.0000	0.0001	0.0003	0.0003	0.0003
	(0.17)	(0.05)	(0.16)	(0.54)	(0.46)	(0.52)
LnBM	0.0025**	0.0024**	0.0022***	0.0023**	0.0022**	0.0021**
	(2.56)	(2.61)	(2.71)	(2.55)	(2.61)	(2.67)
Mom	-0.0021	-0.0020	-0.0019	-0.0013	-0.0013	-0.0012
	(-1.54)	(-1.48)	(-1.46)	(-1.25)	(-1.18)	(-1.13)
Turnover12	-0.0188***	-0.0193***	-0.0203***	-0.0208***	-0.0211***	-0.0220***
	(-2.68)	(-2.95)	(-3.65)	(-3.83)	(-4.10)	(-5.12)
Ivol	-0.0395	-0.0420	-0.0402	-0.0219	-0.0244	-0.0228
	(-1.17)	(-1.30)	(-1.11)	(-0.54)	(-0.63)	(-0.53)
IO	0.0153***	0.0157***	0.0158***	0.0147***	0.0150***	0.0151***
	(6.75)	(6.98)	(7.27)	(6.53)	(6.66)	(6.93)
Constant	-0.0041	-0.0037	-0.0049	-0.0051	-0.0048	-0.0058
	(-1.37)	(-1.32)	(-1.36)	(-1.39)	(-1.36)	(-1.36)
Ave.R-sq	0.051	0.051	0.051	0.050	0.050	0.050
N.of Obs.	121929	121929	121929	121530	121530	121530

- 27% of abnormal return by AIP is concentrated on the three-day earnings window (back to altern)

Double sorts on Firm Size and Abnormal Number of IPs

	Small firms	2	3	4	Large firms
Panel A: Equal-weighted 4 factor alpha					
Low AIP	-0.51	-0.14	-0.27	-0.17	-0.19
2	-0.19	-0.17	-0.22	-0.04	-0.23
3	-0.13	0.09	-0.04	0.01	-0.02
4	0.17	0.11	0.10	0.16	0.20
High AIP	0.64	0.22	0.16	0.20	-0.26
High-Low	**1.14**	**0.36**	**0.43**	**0.37**	-0.07
t-stat	5.38	1.72	2.01	1.68	-0.26
Panel B: Value-weighted 4 factor alpha					
	Small firms	2	3	4	Large firms
Low AIP	-0.57	-0.20	-0.27	-0.19	-0.20
2	-0.28	-0.17	-0.21	-0.04	-0.19
3	-0.15	-0.04	-0.02	-0.01	-0.02
4	-0.03	0.09	0.11	0.15	0.23
High AIP	0.41	0.02	0.19	0.21	-0.30
High-Low	**0.98**	0.22	**0.46**	**0.40**	-0.10
t-stat	4.80	0.97	2.18	1.78	-0.37

Alternative Specifications of AIP

	EW	VW
Model (9) of Expected IP Regression	0.656	0.043
	(3.95)	(0.22)
Nonlinear functional form of Expected IP Regression	0.689	0.552
	(4.30)	(2.39)
Change in AIP relative to 12 months average	0.883	0.388
	(4.82)	(1.44)
Control for lagged # of IPs in Expected IP Regression	0.698	0.508
	(5.44)	(2.03)

- The last two specifications are equivalent to using the innovation in # of IP to predict returns, suggesting the results unlikely explained by (omitted) risk factors

Which Types of EDGAR Filings

	(1)	(2)
AIP_total	-0.0014	-0.0003
	(-0.63)	(-0.17)
AIP_fundl	0.0022	0.0012
	(1.11)	(0.70)
AIP_10K	0.0049***	0.0043***
	(3.96)	(4.02)
Rev		-0.0287***
		(-3.80)
LnME		-0.0014**
		(-2.52)
LnBM		0.0013
		(1.24)
Mom		-0.0048
		(-0.88)
Ivol		-0.0027
		(-0.04)
Turnover12		-0.0088
		(-1.27)
IO		0.0112***
		(3.84)
Constant	0.0122**	0.0120**
	(2.18)	(2.34)
Ave.R-sq	0.005	0.048
N.of Obs.	483667	480793

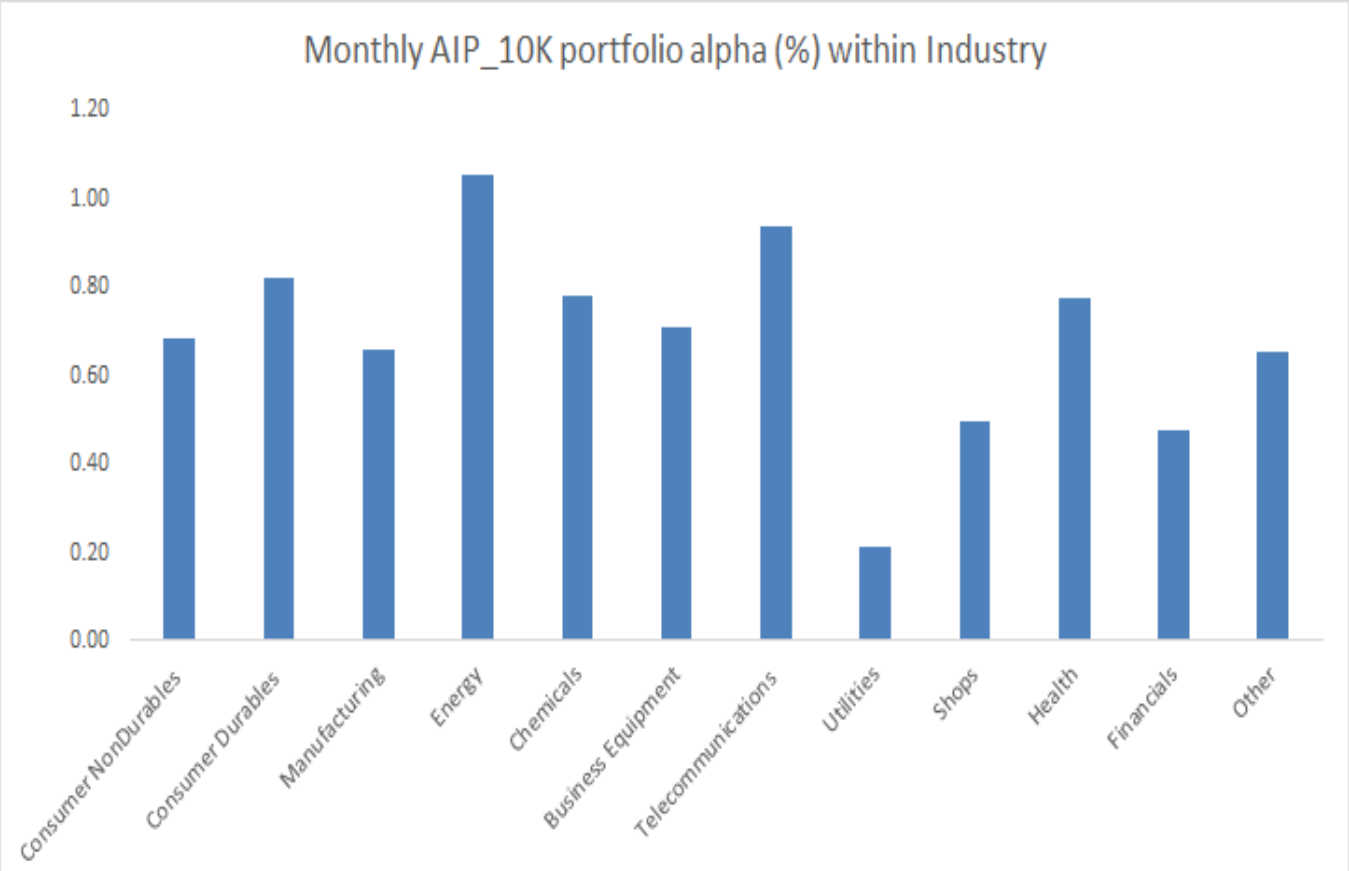
- Return predictability of AIP comes mainly from those searching for firms' annual report 10-K, which are more difficult to analyze and digest compared to other types of financial filings

IPs or Searches?

- Number of searches is dominated by a small fraction of investors who access EDGAR very frequently, and their activities are over-represented (Drake, Roulstone, and Thornock 2015)
- Private information is dispersed among a large group of market participants (Hayek (1945))

	All EDGAR Files		10K, 10Q, 8K		10K	
Asearch	0.0014	-0.0004	0.0020*	-0.0024	0.0033***	-0.0039
	(1.54)	(-0.42)	(1.90)	(-1.49)	(3.93)	(-1.57)
AIP		0.0055**		0.0062***		0.0084***
		(2.45)		(2.83)		(2.90)
Rev	-0.0283***	-0.0284***	-0.0283***	-0.0284***	-0.0284***	-0.0289***
	(-3.73)	(-3.76)	(-3.74)	(-3.77)	(-3.75)	(-3.75)
LnME	-0.0014**	-0.0014***	-0.0014**	-0.0014**	-0.0014***	-0.0013***
	(-2.59)	(-2.63)	(-2.61)	(-2.52)	(-2.64)	(-3.11)
LnBM	0.0013	0.0014	0.0014	0.0014	0.0012	0.0015*
	(1.26)	(1.31)	(1.34)	(1.36)	(1.13)	(1.71)
Mom	-0.0049	-0.0048	-0.0048	-0.0049	-0.0048	-0.0049
	(-0.89)	(-0.88)	(-0.87)	(-0.89)	(-0.86)	(-1.15)
Ivol	0.0048	-0.0014	0.0065	-0.0033	0.0039	-0.0021
	(0.07)	(-0.02)	(0.09)	(-0.05)	(0.05)	(-0.03)
Turnover12	-0.0100	-0.0096	-0.0095	-0.0091	-0.0095	-0.0088
	(-1.46)	(-1.39)	(-1.38)	(-1.33)	(-1.37)	(-1.33)
IO	0.0127***	0.0123***	0.0122***	0.0115***	0.0120***	0.0109***
	(4.10)	(4.04)	(4.06)	(3.86)	(4.03)	(3.57)
Constant	0.0115**	0.0120**	0.0116**	0.0119**	0.0117**	0.0120***
	(2.26)	(2.35)	(2.29)	(2.33)	(2.32)	(3.19)
Ave.R-sq	0.046	0.047	0.046	0.048	0.046	0.049
N.of Obs.	480793	480793	480793	480793	480793	480793

Within-Industry Return Predictability of AIP



- Return predictability of AIP exists in 10 out of 12 Fama-French Industries.

Summary Statistics-Full Sample

Panel A: Summary Statistics

Variable	Mean	Median	STD	P25	P75
Number of IP searching for EDGAR filings					
IP_total	155	94	317	56	159
IP_funtl	107	64	213	37	111
IP_10K	60	32	135	17	60
IP_10Q	37	24	61	13	42
IP_8K	33	19	79	10	36
Stock-level characteristics					
LnME	6.16	6.08	1.98	4.74	7.47
LnBM	-0.66	-0.56	0.84	-1.11	-0.12
Mom	16.67%	7.64%	57.57%	-12.06%	31.78%
Coverage	1.49	1.59	1.01	0.59	2.30
IVOL	0.02	0.02	0.02	0.01	0.03
Turnover12	0.17	0.12	0.19	0.05	0.21
IO	55.30%	59.15%	31.41%	28.92%	80.58%
dROA (%)	0.032	-0.018	4.844	-0.684	0.599
FREV (%)	-0.106	-0.001	22.185	-0.070	0.052
Outflows	-0.10%	-0.05%	0.19%	-0.11%	-0.02%
Lendable Supply	13.96%	14.46%	8.98%	5.85%	20.89%
DCBS	1.48	1.00	1.22	1.00	1.17

- 10-K is the most frequently searched type of SEC filings (back to data)