Third-Party Cookies, Data Sharing, and Return Comovement

Si Cheng[1] Yupeng Lin[2] Ruichang Lu[3] Xiaojun Zhang[3]

[1]Syracuse University
[2]National University of Singapore
[3]Peking University

ABFER 10th Annual Conference, May 2023

Background: The Data Economy

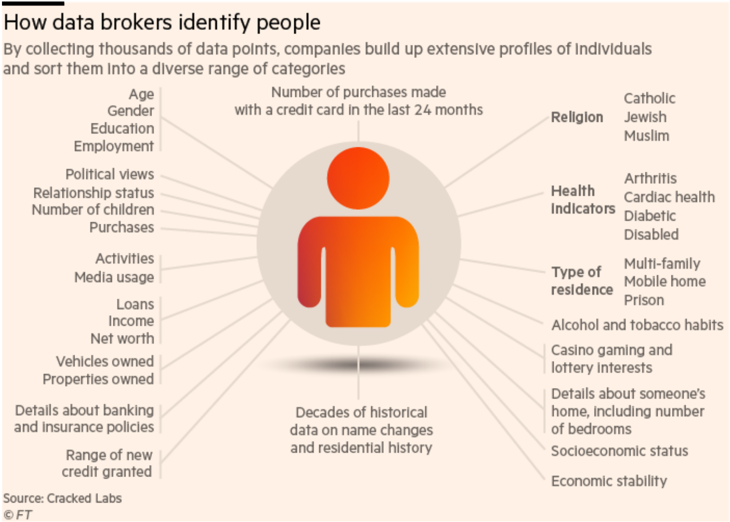
"The world's most valuable resource is no longer oil, but data."

— The Economist (2017)

Background: The Data Economy

- Many companies utilize their digital platforms to collect and commercialize individual data.
- Tracking via tiny files and programs: cookies
  - Enrich the website functionality, e.g., store information (password, address), share content via social networks ▸ Example
  - Allow data brokers to install cookies (third-party cookies) for retargeting and behavioral advertising
  - Track users and their devices across different websites
- The data vendor sales revenue is expected to reach $10.1 billion by 2022, more than tripled from $3.1 billion in 2017.

Background: The Data Economy

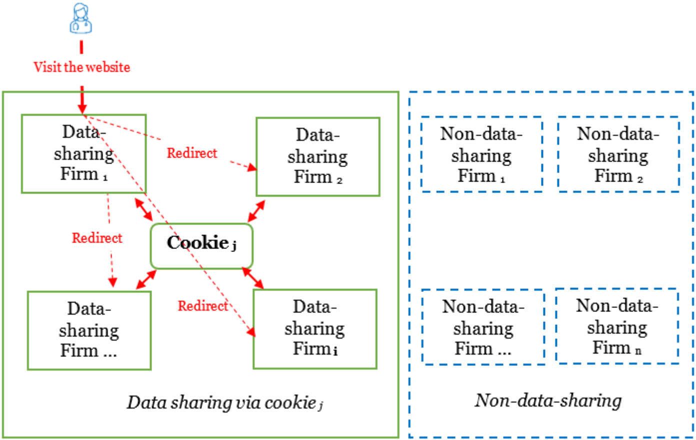


- Display more personalized and targeted ads, based on perceived habits, interests, and recent search activities

Background: Targeted Ads

- The customized popup ads could generate attention shocks for the users, and firms with common third-party cookies are likely to be jointly recommended.

Motivation: Data Sharing via Common Cookies

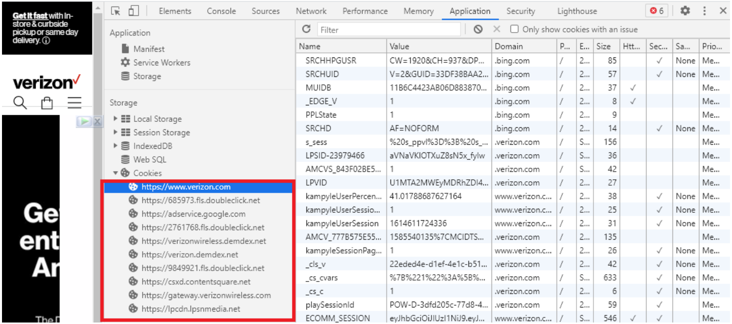


- The same attention shocks could affect investors given the dual nature of investors as consumers.

Research Questions

- Does the data sharing via common cookies affect return comovement?
  - Alleviate the limited attention of investors and help incorporate the new information into stock prices ➔ permanent price adjustment, no reversal
  - Investors overreact to the stale public information and move the price away from fundamentals ➔ temporary price impact, reversal
- What economic mechanism gives rise to the return comovement?
  - Comovement in information acquisition (EDGAR search)
  - Comovement in retail trading

Data on Cookies



- 8/10 are third-party cookies
- 4 cookie platforms: Doubleclick/Google, Demdex/Adobe, Contentsqure, and Ipsnmedia/LivePerson → **cookies** hereafter
- Data-sharing firms: firms with third-party cookies from the same platform

Data

- CRSP: daily and monthly stock characteristics
- COMPUSTAT: quarterly and annual financial statement
- I/B/E/S: analyst forecast
- Thomson-Reuters Institutional Holdings (13F): quarterly institutional equity holdings
- SEC EDGAR Log File: server request records for SEC filings
- TAQ: intraday transactions
- Robintrack: daily number of Robinhood users
- Sample: common stocks on NYSE/AMEX/Nasdaq, 2015–2019
- 4,732 firms, 48% have at least one third-party cookie
- 117 platforms, 9,142 firm-cookie (platform) pairs
- An average firm adopts 4 third-party cookies, with a standard deviation of 5.

Top 10 Cookie Platforms

Platform	No. of Firms with Cookies	No. of Industries with Cookies	Industry Concentration
Doubleclick (Google)	959	62	0.073
Facebook	741	56	0.073
LinkedIn	464	49	0.087
Drawbridge Inc	450	49	0.085
Microsoft	333	47	0.068
Twitter	304	44	0.082
TheTradeDesk	274	46	0.073
Adobe	271	41	0.060
AppNexus Inc	220	43	0.075
Yahoo	141	35	0.078

- The cookie market is dominated by several big platforms.
- The cookie usage is not industry-specific.

Measuring Excess Return Comovement

- Daily regression:

$$\begin{aligned} R_{i,d} &= \alpha_0 + \beta_1 DSRET_{i,d} + \gamma_1 MKT_d \\ &\quad + \gamma_2 SMB_d + \gamma_3 HML_d + \gamma_4 MOM_d + \epsilon_{i,d}, \end{aligned}$$

- $R_{i,d}$: the excess return of stock i on day d
- $DSRET_{i,d}$: the (equal-weighted) excess return of stock i's data-sharing portfolio
- MKT, SMB, HML, MOM: Fama-French-Carhart (FFC) four factors

Baseline Results on Return Comovement: 2015–2019

	Model 1	Model 2	Model 3	Model 4
DSRET	0.266***		0.255***	0.266***
	(15.91)		(15.82)	(15.91)
ResidualDSRET		0.264***		
		(15.87)		
L1.DSRET			0.028***	
			(7.49)	
L2.DSRET			0.004	
			(1.02)	
L3.DSRET			0.007**	
			(2.21)	
L4.DSRET			0.011***	
			(2.77)	
L5_9.DSRET			0.001	
			(0.36)	
FFC Factors	Y	Y	Y	Y
Industry FE	N	N	N	Y

- 1% increase in the data-sharing firms' return → daily abnormal return of 0.27% for the focal firm

Identification Test: CCPA

- California Consumer Privacy Act of 2018 (CCPA)
- Enhance privacy rights and consumer protection for residents of California
  - The right to know about the personal information a business collects about them and how it is used and shared;
  - The right to delete personal information collected from them;
  - The right to opt-out of the sale of their personal information;
  - The right to non-discrimination for exercising their CCPA rights.
- Increase the hurdle of data collection and data sharing via cookies
- The adoption of CCPA is not driven by individual firm characteristics.
- Almost all big firms doing business in California are affected, while the litigation risk is higher for firms with headquarter in California.
- DiD: headquarter in vs. outside California, before vs. after CCPA

Identification Test: CCPA

- California Consumer Privacy Act of 2018 (CCPA)
- Enhance privacy rights and consumer protection for residents of California
  - The right to know about the personal information a business collects about them and how it is used and shared;
  - The right to delete personal information collected from them;
  - The right to opt-out of the sale of their personal information;
  - The right to non-discrimination for exercising their CCPA rights.
- Increase the hurdle of data collection and data sharing via cookies
- The adoption of CCPA is not driven by individual firm characteristics.
- Almost all big firms doing business in California are affected, while the litigation risk is higher for firms with headquarter in California.
- DiD: headquarter in vs. outside California, before vs. after CCPA

Identification Test: CCPA

	Model 1	Model 2	Model 4	Model 5
DSRET_CA	0.312***	0.403***	0.312***	0.405***
	(8.03)	(8.90)	(8.02)	(8.94)
DSRET_non-CA	0.422***	0.418***	0.422***	0.418***
	(16.91)	(16.64)	(16.93)	(16.65)
DSRET_CA × Post 2Y		-0.165***		-0.167***
		(-2.62)		(-2.66)
DSRET_non-CA × Post 2Y		0.010		0.011
		(0.62)		(0.67)
Post	N	Y	N	Y
FFC Factors	Y	Y	Y	Y
Industry FE	N	N	Y	Y

- After the introduction of CCPA, focal firm:
  - Comoves less with California-data-sharing firms ➔ 41% decline
  - No change in comovement with non-California-data-sharing firms
- $DiD = -0.175^{**}$

Identification Test: CCPA

	Model 3	Model 6
DSRET_CA	0.406***	0.407***
	(8.97)	(9.01)
DSRET_non-CA	0.418***	0.418***
	(16.77)	(16.79)
DSRET_CA $\times$ Post^{+1}	-0.082	-0.085
	(-1.03)	(-1.07)
DSRET_non-CA $\times$ Post^{+1}	-0.007	-0.006
	(-0.33)	(-0.30)
DSRET_CA $\times$ Post^{+2}	-0.234***	-0.237***
	(-2.92)	(-2.94)
DSRET_non-CA $\times$ Post^{+2}	0.021	0.022
	(1.07)	(1.11)
Post	Y	Y
FFC Factors	Y	Y
Industry FE	N	Y

- Stronger results as effective date (Jan 1, 2020) draws closer
- $DiD = -0.075$ for Year 1 and -0.255^{***} for Year 2

Robustness Check: Pairwise Return Comovement

- Monthly Fama-MacBeth regression:

$$ARCORR_{ij,t} = \alpha_0 + \beta_1 DS_{ij,t-1} + \boldsymbol{\gamma}' \boldsymbol{N}_{ij,t-1} + \epsilon_{ij,t},$$

- $ARCORR_{ij,t}$: the correlation of daily FFC four-factor abnormal returns between stocks i and j in month t
- $DS_{ij,t-1}$: data sharing variables
  - $NUMTP$: the number of common third-party cookies
  - $LOGNUMTP$: the logarithm of $(1 + NUMTP)$
  - $\%NUMTP$: the percentage of common third-party cookies
  - $DNUMTP$: a dummy variable indicating $NUMTP > 0$
- $\boldsymbol{N}$: 30 control variables following Antón and Polk (2014)
  - Common institutional ownership and analyst coverage
  - Similarity in firm characteristics (size, book-to-market, momentum, past return, return on equity, abnormal trading volume, sales growth, financial leverage, share price, state, index affiliation, and stock exchange) and industry classification
  - Interactions and higher moments

Robustness Check: Pairwise Return Comovement

	Model 4	Model 6	Model 8	Model 10
NUMTP	0.125***			
	(8.33)			
LOGNUMTP		0.358***		
		(7.83)		
%NUMTP			1.416***	
			(7.63)	
DNUMTP				0.279***
				(6.88)
Pair Controls	Y	Y	Y	Y
Style Controls	Y	Y	Y	Y

- 1 std.dev. increase in *NUMTP* → 0.19% higher residual return comovement (52% increase relative to the sample average)
- Stock pairs with common cookies → 0.28% higher residual return comovement (76% increase relative to the sample average)

Comovement in Information Acquisition

- Explicitly measure information acquisition based on EDGAR search
- Daily regression:

$$\begin{aligned} HUM_{i,d} &= \alpha_0 + \beta_1 DSHUM_{i,d} + \beta_2 DSROB_{i,d} \\ &\quad + \gamma_1 MKTHUM_d + \gamma_2 MKTROB_d + \epsilon_{i,d}, \end{aligned}$$

- $HUM_{i,d}$: the number of human search of stock i on day d

	Model 3	Model 4
DSHUM	0.731***	0.707***
	(22.32)	(14.51)
DSROB		0.006
		(0.86)
MKTHUM	0.302***	0.315***
	(7.72)	(5.57)
MKTROB		-0.004
		(-0.46)
Industry FE	Y	Y

Comovement in Information Acquisition

- Explicitly measure information acquisition based on EDGAR search
- Daily regression:

$$\begin{aligned} HUM_{i,d} &= \alpha_0 + \beta_1 DSHUM_{i,d} + \beta_2 DSROB_{i,d} \\ &\quad + \gamma_1 MKTHUM_d + \gamma_2 MKTROB_d + \epsilon_{i,d}, \end{aligned}$$

- $HUM_{i,d}$: the number of human search of stock i on day d

	Model 3	Model 4
DSHUM	0.731***	0.707***
	(22.32)	(14.51)
DSROB		0.006
		(0.86)
MKTHUM	0.302***	0.315***
	(7.72)	(5.57)
MKTROB		-0.004
		(-0.46)
Industry FE	Y	Y

Comovement in Retail Trading

- Proxies for retail trading:
  - Order imbalance for retail trades based on share volume: *OIBVOL* (Boehmer et al. 2021)
  - Change and percentage change in the number of Robinhood users: *RHNUM* and *RHPCT* (Barber et al. 2022)
- Daily regression: $TRA_{i,d} = \alpha_0 + \beta_1 DSTRA_{i,d} + \gamma_1 MKTTRA_d + \epsilon_{i,d}$,

	OIBVOL Model 3	OIBVOL Model 4	RHNUM Model 6	RHPCT Model 8
DSOIBVOL	0.088***			
	(8.25)			
$DSOIBVOL^+$		0.130***		
		(6.05)		
$DSOIBVOL^-$		0.053***		
		(3.03)		
DSRHNUM			0.372***	
			(5.15)	
DSRHPCT				0.084*
				(1.80)
MKTTRA	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y

Comovement in Retail Trading

- Proxies for retail trading:
  - Order imbalance for retail trades based on share volume: *OIBVOL* (Boehmer et al. 2021)
  - Change and percentage change in the number of Robinhood users: *RHNUM* and *RHPCT* (Barber et al. 2022)
- Daily regression: $TRA_{i,d} = \alpha_0 + \beta_1 DSTRA_{i,d} + \gamma_1 MKTTRA_d + \epsilon_{i,d}$,

	OIBVOL Model 3	OIBVOL Model 4	RHNUM Model 6	RHPCT Model 8
DSOIBVOL	0.088***			
	(8.25)			
DSOIBVOL^{+}		0.130***		
		(6.05)		
DSOIBVOL^{-}		0.053***		
		(3.03)		
DSRHNUM			0.372***	
			(5.15)	
DSRHPCT				0.084*
				(1.80)
MKTTRA	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y

Link Correlated Investor Behavior to Return Comovement

- Monthly Fama-MacBeth regression:

$$\begin{aligned} ARCORR_{ij,t} &= \alpha_0 + \beta_1 DS_{ij,t-1} + \beta_2 DS_{ij,t-1} \times IBCORR_{ij,t-1} \\ &+ \beta_3 IBCORR_{ij,t-1} + \gamma' \mathbf{N}_{ij,t-1} + \epsilon_{ij,t}, \end{aligned}$$

	Model 2	Model 4	Model 6	Model 8	Model 10
NUMTP	0.078**	0.134***	0.123***	0.125***	0.124***
	(2.73)	(8.45)	(6.74)	(5.49)	(5.58)
NUMTP × HUMCORR	0.115***				
	(2.81)				
NUMTP × ROBCORR	-0.017				
	(-0.68)				
NUMTP × OIBVOLCORR		0.031**			
		(2.30)			
NUMTP × POSOIBVOL			0.091**		
			(2.06)		
NUMTP × NEGOIBVOL			-0.050		
			(-1.47)		
NUMTP × RHNUMCORR				0.063***	
				(3.45)	
NUMTP × RHPCTCORR					0.065***
					(3.53)
Pair Controls	Y	Y	Y	Y	Y
Style Controls	Y	Y	Y	Y	Y

Additional Analyses

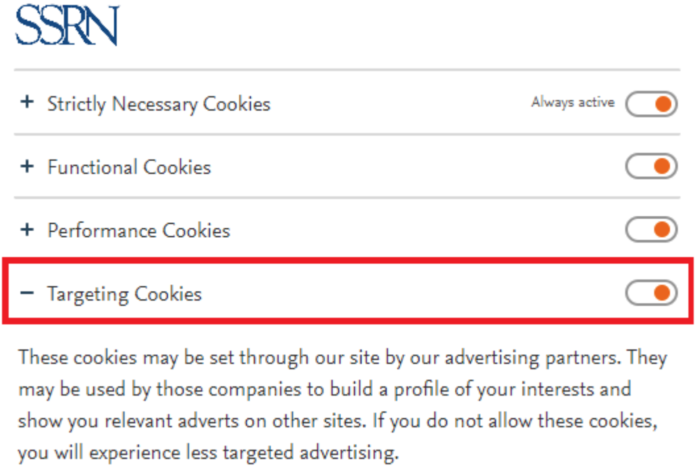
- Cross-sectional variation in return comovement
  - Consumer-related vs. other industries ▸ Details
  - High-frequency vs. low-frequency cookies ▸ Details
- Trading strategy ▸ Details
  - Long stocks with high data-sharing portfolio returns but low own returns
  - Short stocks with low data-sharing portfolio returns but high own returns
- Cash flow comovement in the long-term ▸ Details
  - No significant comovement in advertising expenditure, sales, and R&D expense

Conclusion

- Capital market consequences of online data sharing
  1. More joint search for financial information and correlated retail buying
  2. Higher return comovement, especially for consumer-related industries and frequently installed cookies
  3. Enhanced information diffusion
- Beneficial effect of data sharing: alleviates the limited attention of investors and improves the price efficiency
- Unique role compared to traditional information intermediaries and distribution methods

Appendix

Background: The Data Economy



Back

Cross-Sectional Variation in Return Comovement

- Subsamples by industry classification ▸ Details

	Consumer-Related Industries			Other Industries		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
DSRET	0.435***		0.435***	0.180***		0.180***
	(20.22)		(20.29)	(10.19)		(10.16)
ResidualDSRET		0.418***			0.185***	
		(19.69)			(10.40)	
FFC Factors	Y	Y	Y	Y	Y	Y
Industry FE	N	N	Y	N	N	Y

- 1% increase in the data-sharing firms' return → daily abnormal return of 0.44% (0.18%) for the focal firm in consumer-related (other) industries

◂ Back

Cross-Sectional Variation in Return Comovement

- Subsamples by cookie adoption

	High-Frequency Cookies			Low-Frequency Cookies		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
DSRET	0.551***		0.551***	0.175***		0.175***
	(24.29)		(24.29)	(17.17)		(17.13)
ResidualDSRET		0.540***			0.174***	
		(23.97)			(17.14)	
FFC Factors	Y	Y	Y	Y	Y	Y
Industry FE	N	N	Y	N	N	Y

- 1% increase in the data-sharing firms' return → daily abnormal return of 0.55% (0.18%) for the focal firm with high-frequency (low-frequency) cookies

Back

Consumer-Related Industries

- Firms with two-digit SIC code of
  - 01-09 Agriculture, Forestry, and Fishing
  - 40-49 Transportation, Communications, Electric, Gas, and Sanitary Services
  - 52-59 Retail Trade
  - 60-67 Finance, Insurance, and Real Estate
  - 70-89 Services

Back

Trading Strategy

- Can we exploit the lead-lag return predictability induced by the information diffusion?
- Suppose stocks i and j share common third-party cookies.
  - Stock i: focal stock
  - Stock j: stock i's data-sharing stock
- If stock i's price (i.e., own price) declines while data-sharing stock j's price increases ➔ stock i is relatively undervalued and its price should go up once investors learn about the firm
- Independently sort stocks into quintile portfolios according to their own returns and data-sharing portfolio returns (weighted by the number of common third-party cookies)
- Risk-adjusted return from a five-factor model: FFC four factors and the short-term reversal factor

Trading Strategy: 1-Day Holding Period, 5-Factor Alpha

Own Return	Data-Sharing Portfolio Return: Low	2	3	4	High	HML	
Low	0.098***	0.111***	0.115***	0.144***	0.125***	0.027	
	(5.48)	(6.57)	(7.56)	(9.19)	(7.32)	(1.45)	
2	0.004	0.023**	0.021***	0.020**	0.005	0.001	
	(0.40)	(2.53)	(2.64)	(2.22)	(0.52)	(0.13)	
3	0.002	0.008	-0.004	0.005	0.015**	0.013	
	(0.27)	(1.12)	(-0.56)	(0.56)	(2.06)	(1.13)	
4	-0.015*	-0.009	-0.004	0.000	-0.006	0.010	
	(-1.91)	(-1.05)	(-0.54)	(0.03)	(-0.66)	(0.95)	
High	-0.144***	-0.126***	-0.100***	-0.099***	-0.084***	0.059***	
	(-8.18)	(-7.68)	(-7.22)	(-6.31)	(-4.67)	(3.22)	
LMH	0.242***	0.237***	0.215***	0.243***	0.209***	0.268***	HL − LH
	(10.10)	(10.36)	(10.23)	(10.91)	(8.33)	(11.88)	

- The long-short portfolio yields a daily five factor-adjusted return of 0.27%.

Trading Strategy: 5-Day Holding Period, 5-Factor Alpha

Own Return	Data-Sharing Portfolio Return: Low	2	3	4	High	HML
Low	0.018	0.029***	0.030***	0.039***	0.029**	0.011
	(1.44)	(3.02)	(3.12)	(4.01)	(2.45)	(1.33)
2	0.000	0.011**	0.010**	0.013***	0.003	0.002
	(0.04)	(2.22)	(1.97)	(2.66)	(0.43)	(0.47)
3	0.007	0.005	0.003	0.010**	0.007	-0.000
	(1.46)	(1.26)	(0.81)	(2.20)	(1.39)	(-0.09)
4	-0.005	0.007	0.003	0.008	0.006	0.010**
	(-0.88)	(1.41)	(0.69)	(1.64)	(1.11)	(2.18)
High	-0.032***	-0.028***	-0.021**	-0.013	-0.019*	0.013*
	(-2.91)	(-3.10)	(-2.43)	(-1.53)	(-1.85)	(1.70)
LMH	0.050***	0.058***	0.051***	0.053***	0.048***	0.061*** HL − LH
	(4.80)	(6.18)	(5.18)	(5.33)	(4.57)	(6.22)

- The return predictability is short-lived → 88% $(0.268/(0.061 \times 5))$ of the 5-day risk-adjusted return is concentrated on the first day.

Trading Strategy: 21-Day Holding Period, 5-Factor Alpha

Own Return	Data-Sharing Portfolio Return Low	2	3	4	High	HML
Low	0.009	0.012	0.014*	0.016**	0.012	0.003
	(0.76)	(1.45)	(1.71)	(1.97)	(1.15)	(0.80)
2	0.003	0.009**	0.010***	0.007*	0.005	0.002
	(0.74)	(2.51)	(2.58)	(1.84)	(1.07)	(0.64)
3	0.006	0.012***	0.010***	0.011***	0.007*	0.001
	(1.59)	(3.38)	(2.82)	(3.08)	(1.70)	(0.24)
4	0.002	0.005	0.006	0.007**	0.005	0.003
	(0.45)	(1.40)	(1.52)	(2.00)	(1.22)	(1.24)
High	-0.013	-0.007	-0.002	-0.000	-0.010	0.003
	(-1.40)	(-0.86)	(-0.25)	(-0.06)	(-1.07)	(0.81)
LMH	0.022***	0.018***	0.016***	0.017***	0.022***	0.025*** HL − LH
	(3.93)	(4.18)	(3.64)	(3.59)	(4.28)	(5.05)

- The return predictability is short-lived → 51% of the 21-day risk-adjusted return is concentrated on the first day.

Trading Strategy: 22–252-Day Holding Period, 5-Factor Alpha

Own Return	Data-Sharing Portfolio Return Low	2	3	4	High	HML	
Low	-0.005	0.003	0.005	0.004	-0.005	0.000	
	(-0.50)	(0.38)	(0.62)	(0.49)	(-0.52)	(0.06)	
2	0.002	0.008	0.010*	0.006	0.002	-0.000	
	(0.31)	(1.47)	(1.92)	(1.24)	(0.31)	(-0.00)	
3	0.003	0.009*	0.010**	0.008*	0.003	-0.000	
	(0.56)	(1.72)	(2.09)	(1.65)	(0.56)	(-0.01)	
4	0.002	0.006	0.010**	0.005	0.001	-0.001	
	(0.28)	(1.22)	(2.09)	(1.09)	(0.19)	(-0.48)	
High	-0.004	0.003	0.005	0.001	-0.006	-0.002	
	(-0.46)	(0.34)	(0.62)	(0.18)	(-0.65)	(-0.77)	
LMH	-0.001	0.000	0.001	0.003	0.000	-0.001	HL − LH
	(-0.38)	(0.22)	(0.32)	(1.21)	(0.15)	(-0.55)	

- No subsequent reversal up to one year → permanent price adjustment

Back

Cash Flow Comovement in the Long-Term

- Annual regression:

$$Char_{i,y} = \alpha_0 + \beta_1 DSChar_{i,y} + \boldsymbol{\gamma}' \boldsymbol{C}_{i,y} + \epsilon_{i,y},$$

- $\boldsymbol{C}$: market-to-book, tangibility, total assets, current ratio, coverage ratio, Z-score, and leverage

	Advertising Model 1	Sales Model 2	R&D Model 3	Advertising Model 4	Sales Model 5	R&D Model 6
DSAdv	0.454*			0.133		
	(1.84)			(0.53)		
DSSALE		0.095**			0.033	
		(2.22)			(0.85)	
DSRD			0.162			-0.021
			(0.83)			(-0.11)
Firm Controls	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Industry FE	N	N	N	Y	Y	Y

Back