# The effect of Google search volume on retail investor trades

**Matthias Bank, Felix Kunz, Alexander Kupfer and Markus G. Schmidt**

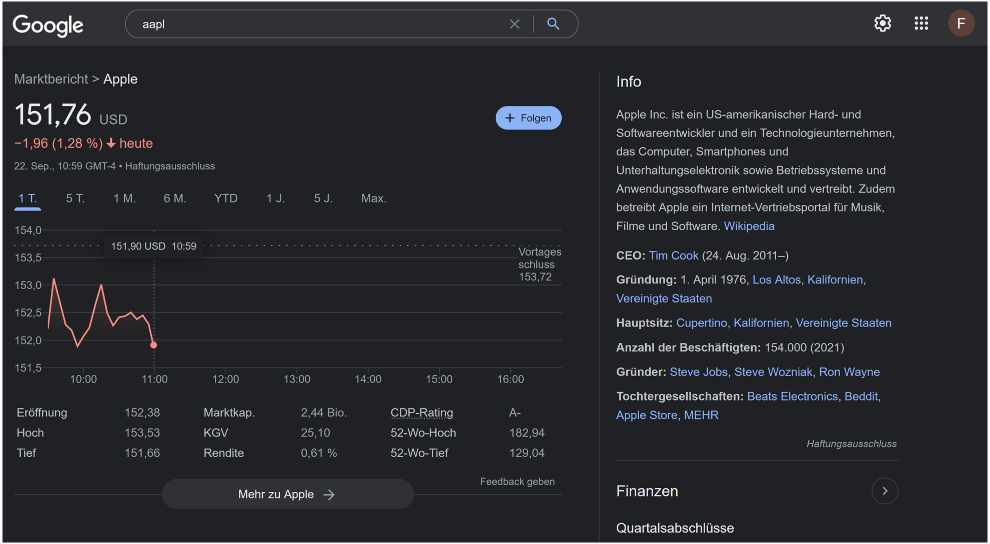
*University of Innsbruck*

37. Workshop of the Austrian Working Group on Banking and Finance

September 23, 2022

# Which type of investor knows this screen?

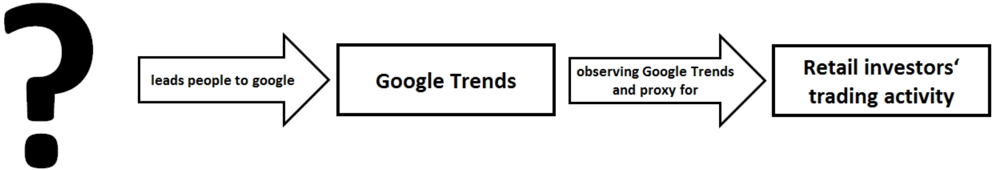
# … and which investor knows this screen?



# Do retail investors use Google and trade thereafter?

Starting with Da, Engelberg, et al. (2011) and their cautious statement that Google's Search Volume Index (SVI) "likely measures the attention of retail investors":

- assumed that retail investors use Google to inform themselves about stocks
- consensus that Google Trends captures the attention of retail investors



# SVI implicitly assumed as retail investor attention

- **Stock returns**
  Joseph et al. (2011); Bae and W. Wang (2012); Takeda and Wakao (2014); Zhang and Y. Wang (2015); Mbanga et al. (2019); Hao and Xiong (2021); Yuan et al. (2022); Da, Hua, et al. (2022)
- **Fixed Income, Commodities, Futures and REIT returns**
  Vozlyublennaia (2014); Chen et al. (2016); Yung and Nafar (2017)
- **Volatility**
  Hamid and Heiden (2015); Irresberger et al. (2015); Dimpfl and Jank (2016)

# SVI implicitly assumed as retail investor attention – cont'd

- **Liquidity**
  Da, Engelberg, et al. (2011); Bank et al. (2011); Ding and Hou (2015); Cheng et al. (2021)
- **IPO valuation**
  Colaco et al. (2017); Kao et al. (2022)
- **Herding**
  Hsieh et al. (2020); Gavish et al. (2021); Wanidwaranan and Padungsaksawasdi (2022)

# Contribution

- Combination of SVI and orderbook data
- Observing effects of Abnormal search volume (ASVI) on retail investor trades
- **New approach:** directly observing retail investor trades, based on the trade identification algorithm by Boehmer et al. (2021)

**Finding I**

- ASVI positively related to retail investor trading activity

**Finding II**

- ASVI positively related to all other market participants trading activity as well

# S&P 500 constituents

- All 550 constituents of the S&P 500 during the years 2018 and 2019 from CRSP

- Orderbook data (NBBO and Trades) from NYSE TAQ
- Exchanges: NYSE, AMEX, NASDAQ, NYSE Arca
- Sample period of 502 trading days: 01.01.2018 until 31.12.2019

- News data from Bloomberg
- Price and return data from CRSP

- Search volume index with geographical location US from Google Trends

# Google search volume index [1]

Table: SVI – Screening procedure

| | |
|---|---|
| All Ticker | 550 |
| Excluded in Step 1<br>*Ticker is a word* | 18 |
| Excluded in Step 2<br>*No financial information supplied* | 101 |
| Excluded in Step 3<br>*No financial information demanded* | 48 |
| Excluded in Step 4<br>*At least 50% of the time SVI is available* | 85 |
| Valid Ticker | 298 |

[1] following Niessner (2015) and Kupfer and Schmidt (2021)

# Retail investor trades

We apply the trade classification algorithm by **Boehmer et al. (2021)**

- most retail trades occur off-exchange via wholesalers or internalisation
- appear as trades with exchange code "D" in TAQ data
- buyer-initiated trades by retail investors are below a round penny (sub-penny interval of [0.6, 1])
- seller-initiated trades by retail investors are above a round penny (sub-penny interval of [0, 0.4])

# Remaining trades

We apply the three most popular trade classification algorithms and match NBBO and Trade by up to one millisecond, following Holden and Jacobsen (2014)

**Lee and Ready (1991)** convention

- buyer initiated if $Price_i > Midpoint_i$
- seller initiated if $Price_i < Midpoint_i$

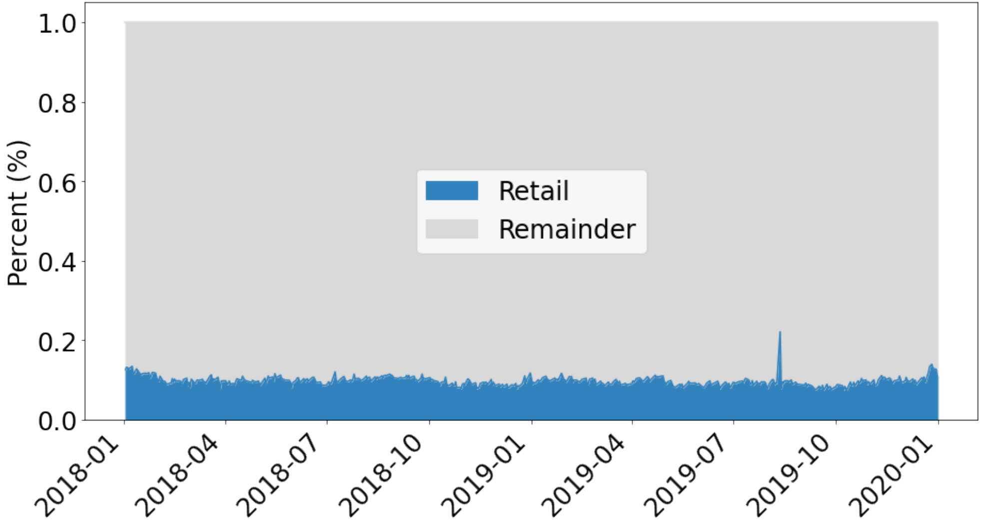
**Ellis et al. (2000)** convention

- buyer initiated if $Price_i = Ask_i$
- seller initiated if $Price_i = Bid_i$

**Chakrabarty et al. (2007)** convention

- buyer initiated if $Price_i \in [0.3 * Bid_i + 0.7 * Ask_i, Ask_i]$
- seller initiated if $Price_i \in [Bid_i, 0.7 * Bid_i + 0.3 * Ask_i]$

# Total dollar volume of buyer initiated trades



# ASVI, abnormal buys and sells and other controls

**Abnormal search volume index**

$ASVI_{i,t} = ln(SVI_{i,t}) - ln(Mean(SVI_{i,t-1}, ..., SVI_{i,t-20}))$

**Abnormal number of buyer (seller) initiated trades**

$ANBuys_{i,t} = ln(NBuys_{i,t}) - ln(Mean(NBuys_{i,t-1}, ..., NBuys_{i,t-20}))$

Other investor attention proxies as used in Ben-Rephael et al. (2017)

- Abnormal news: $ANews_{i,t} = ln(1 + News_{i,t}) - ln(Mean(News_{i,t-1}, ..., News_{i,t-20}))$
- Abnormal trading volume: $AVol_{i,t} = ln(Vol_{i,t}) - ln(Mean(Vol_{i,t-1}, ..., Vol_{i,t-20}))$
- $Ret_{i,t}$, $Abs(Ret_{i,t})$, $Ret_{i,t-1:t-5}$ and $Ret_{i,t-6:t-53}$ – Actual and past returns
- $\sigma_{i,t-1:t-5}$ – Past volatility
- $Spread_{i,t}$ – Percentage effective spread
- $IntraVola_{i,t}$ – Intraday volatility
- $HLtoH_{i,t}$ – Relative price range

# Abnormal number of buyer initiated trades

$$AN_{i,t}^{Buys,c} = \alpha + \beta * ASVI_{i,t} + \sum_{k=1}^{K} \theta_k * X_{i,t}^{k} + \tau_t + \mu_i + \epsilon_{i,t} \quad (1)$$

| | (1) Retail | (2) LR | (3) EMO | (4) CLNV |
|---|---|---|---|---|
| *ASVI* | 0.043*** | 0.018*** | 0.018*** | 0.018*** |
| | (0.002) | (0.001) | (0.001) | (0.001) |
| *ANews* | 0.055*** | 0.039*** | 0.039*** | 0.039*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $Avol_{t-1}$ | 0.301*** | 0.343*** | 0.336*** | 0.338*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| *Other Controls* | Yes | Yes | Yes | Yes |
| *Time FE* | Yes | Yes | Yes | Yes |
| *Firm FE* | Yes | Yes | Yes | Yes |
| $R^2$ | 0.456 | 0.604 | 0.630 | 0.615 |
| *Observations* | 121,267 | 121,267 | 121,267 | 121,267 |

Complete regression output: ANBuys

# Abnormal number of seller initiated trades

$$AN^{Sells,c}_{i,t} = \alpha + \beta * ASVI_{i,t} + \sum_{k=1}^{K} \theta_k * X^k_{i,t} + \tau_t + \mu_i + \epsilon_{i,t} \quad (2)$$

| | (1) Retail | (2) LR | (3) EMO | (4) CLNV |
|---|---|---|---|---|
| *ASVI* | 0.042*** | 0.019*** | 0.019*** | 0.019*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| *ANews* | 0.048*** | 0.043*** | 0.043*** | 0.043*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $Avol_{t-1}$ | 0.271*** | 0.349*** | 0.364*** | 0.356*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| *Other Controls* | Yes | Yes | Yes | Yes |
| *Time FE* | Yes | Yes | Yes | Yes |
| *Firm FE* | Yes | Yes | Yes | Yes |
| $R^2$ | 0.513 | 0.619 | 0.603 | 0.612 |
| *Observations* | 121,267 | 121,267 | 121,267 | 121,267 |

Complete regression output: ANSells

# Beware when using Google!

The ‘implicit’ assumption that Google Trends captures the attention of retail investors seems to be true, but there is some by-catch.

Our results indicate that:

- Google Trends captures an unobserved attention generating event
- those events lead retail investors to trade
- **but** those events also lead all other market participants to trade

# References I


- Bae, Kee-Hong and Wei Wang (2012). "What's in a "China" name? A test of investor attention hypothesis". In: *Financial Management* 41.2, pp. 429–455.
- Bank, Matthias, Martin Larch, and Georg Peter (2011). "Google search volume and its influence on liquidity and returns of German stocks". In: *Financial Markets and Portfolio Management* 25.3, pp. 239–264.
- Ben-Rephael, Azi, Zhi Da, and Ryan D Israelsen (2017). "It Depends on Where You Search: Institutional Investor Attention and Underreaction to News". In: *The Review of Financial Studies* 30.9, pp. 3009–3047.
- Boehmer, Ekkehart et al. (2021). "Tracking retail investor activity". In: *The Journal of Finance* 76.5, pp. 2249–2305.
- Chakrabarty, Bidisha et al. (2007). "Trade classification algorithms for electronic communications network trades". In: *Journal of Banking & Finance* 31.12, pp. 3806–3821.

# References II


- Chen, Jing et al. (2016). "Investor attention and macroeconomic news announcements: Evidence from stock index futures". In: *Journal of Futures Markets* 36.3, pp. 240–266.
- Cheng, Feiyang et al. (2021). "Does retail investor attention improve stock liquidity? A dynamic perspective". In: *Economic Modelling* 94, pp. 170–183.
- Colaco, Hugh MJ, Amedeo De Cesari, and Shantaram P Hegde (2017). "Retail investor attention and IPO valuation". In: *European Financial Management* 23.4, pp. 691–727.
- Da, Zhi, Joseph Engelberg, and Pengjie Gao (2011). "In Search of Attention". In: *The Journal of Finance* 66.5, pp. 1461–1499.
- Da, Zhi, Jian Hua, et al. (2022). "Market Returns and a Tale of Two Types of Attention". In: *Available at SSRN 3551662*.

# References III


- Dimpfl, Thomas and Stephan Jank (2016). "Can internet search queries help to predict stock market volatility?" In: *European Financial Management* 22.2, pp. 171–192.
- Ding, Rong and Wenxuan Hou (2015). "Retail investor attention and stock liquidity". In: *Journal of International Financial Markets, Institutions and Money* 37, pp. 12–26.
- Ellis, Katrina, Roni Michaely, and Maureen O'Hara (2000). "The accuracy of trade classification rules: Evidence from Nasdaq". In: *Journal of Financial and Quantitative Analysis* 35.4, pp. 529–551.
- Gavish, Liron Reiter, Mahmoud Qadan, and Joseph Yagil (2021). "Net buyers of attention-grabbing stocks? Who exactly are they?" In: *Journal of Behavioral Finance* 22.1, pp. 26–45.

# References IV


- Hamid, Alain and Moritz Heiden (2015). "Forecasting volatility with empirical similarity and Google Trends". In: *Journal of Economic Behavior & Organization* 117, pp. 62–81.
- Hao, Jing and Xiong Xiong (2021). "Retail investor attention and firms' idiosyncratic risk: Evidence from China". In: *International Review of Financial Analysis* 74, p. 101675.
- Holden, Craig W and Stacey Jacobsen (2014). "Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions". In: *The Journal of Finance* 69.4, pp. 1747–1785.
- Hsieh, Shu-Fan, Chia-Ying Chan, and Ming-Chun Wang (2020). "Retail investor attention and herding behavior". In: *Journal of Empirical Finance* 59, pp. 109–132.
- Irresberger, Felix, Janina Mühlnickel, and Gregor NF Weiß (2015). "Explaining bank stock performance with crisis sentiment". In: *Journal of Banking & Finance* 59, pp. 311–329.

# References V


- Joseph, Kissan, M Babajide Wintoki, and Zelin Zhang (2011). "Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search". In: *International Journal of Forecasting* 27.4, pp. 1116–1127.
- Kao, Lanfeng, Anlin Chen, and Cheng-Shou Lu (2022). "Retail investor attention and IPO prices with a pre-IPO market". In: *International Review of Economics & Finance* 82, pp. 416–432.
- Kupfer, Alexander and Markus G Schmidt (2021). "In search of retail investors: The effect of retail investor attention on odd lot trades". In: *Journal of Empirical Finance* 62, pp. 315–326.
- Lee, Charles MC and Mark J Ready (1991). "Inferring trade direction from intraday data". In: *The Journal of Finance* 46.2, pp. 733–746.
- Mbanga, Cedric, Ali F Darrat, and Jung Chul Park (2019). "Investor sentiment and aggregate stock returns: the role of investor attention". In: *Review of Quantitative Finance and Accounting* 53.2, pp. 397–428.

# References VI


- Niessner, Marina (2015). "Strategic disclosure timing and insider trading". In: *Available at SSRN 2439040*.
- Takeda, Fumiko and Takumi Wakao (2014). "Google search intensity and its relationship with returns and trading volume of Japanese stocks". In: *Pacific-Basin Finance Journal* 27, pp. 1–18.
- Vozlyublennaia, Nadia (2014). "Investor attention, index performance, and return predictability". In: *Journal of Banking & Finance* 41, pp. 17–35.
- Wanidwaranan, Phasin and Chaiyuth Padungsaksawasdi (2022). "Unintentional herd behavior via the Google search volume index in international equity markets". In: *Journal of International Financial Markets, Institutions and Money* 77, p. 101503.
- Yuan, Ying, Xiaoqian Fan, and Yiou Li (2022). "Do local and non-local retail investor attention impact stock returns differently?" In: *Pacific-Basin Finance Journal* 74, p. 101807.

# References VII


Yung, Kenneth and Nadia Nafar (2017). "Investor attention and the expected returns of reits". In: *International Review of Economics & Finance* 48, pp. 423–439.

Zhang, Bing and Yudong Wang (2015). "Limited attention of individual investors and stock performance: Evidence from the ChiNext market". In: *Economic Modelling* 50, pp. 94–104.

# *ANBuys* – Complete regression output

| | (1) Retail | (2) LR | (3) EMO | (4) CLNV |
|---|---|---|---|---|
| *Constant* | -0.451*** | -0.199*** | -0.186*** | -0.194*** |
| | (0.018) | (0.015) | (0.014) | (0.014) |
| *ASVI* | 0.043*** | 0.018*** | 0.018*** | 0.018*** |
| | (0.002) | (0.001) | (0.001) | (0.001) |
| *ANews* | 0.055*** | 0.039*** | 0.039*** | 0.039*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $Avol_{t-1}$ | 0.301*** | 0.343*** | 0.336*** | 0.338*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| *HLtoH* | 3.542*** | 3.970*** | 4.128*** | 4.037*** |
| | (0.139) | (0.114) | (0.108) | (0.111) |
| *Spread* | -23.818*** | -87.652*** | -82.300*** | -87.018*** |
| | (3.521) | (2.873) | (2.738) | (2.804) |
| *IntraVola* | 236.885*** | 292.304*** | 296.848*** | 293.664*** |
| | (5.315) | (4.337) | (4.134) | (4.233) |
| $Ret_t$ | -0.740*** | 0.479*** | 0.242*** | 0.390*** |
| | (0.052) | (0.042) | (0.040) | (0.041) |
| $Abs(Ret_t)$ | 5.904*** | 4.051*** | 3.786*** | 3.920*** |
| | (0.088) | (0.071) | (0.068) | (0.070) |
| $Ret_{t-1:t-5}$ | -0.402*** | -0.171*** | -0.159*** | -0.165*** |
| | (0.024) | (0.019) | (0.018) | (0.019) |
| $Ret_{t-6:t-53}$ | 0.196*** | 0.070*** | 0.095*** | 0.080*** |
| | (0.008) | (0.007) | (0.007) | (0.007) |
| $\sigma_{t-1:t-5}$ | -2.942*** | -3.104*** | -3.133*** | -3.102*** |
| | (0.098) | (0.080) | (0.077) | (0.078) |
| *52HighDum* | 0.095*** | 0.047*** | 0.040*** | 0.043*** |
| | (0.004) | (0.003) | (0.003) | (0.003) |
| *52LowDum* | 0.097*** | 0.032*** | 0.026*** | 0.030*** |
| | (0.006) | (0.005) | (0.005) | (0.005) |
| *Time FE* | Yes | Yes | Yes | Yes |
| *Firm FE* | Yes | Yes | Yes | Yes |
| $R^2$ | 0.456 | 0.604 | 0.630 | 0.615 |
| *Observations* | 121,267 | 121,267 | 121,267 | 121,267 |

# *ANSells* – Complete regression output

| | (1) Retail | (2) LR | (3) EMO | (4) CLNV |
|---|---|---|---|---|
| *Constant* | -0.444*** | -0.194*** | -0.209*** | -0.199*** |
| | (0.015) | (0.015) | (0.015) | (0.015) |
| *ASVI* | 0.042*** | 0.019*** | 0.019*** | 0.019*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| *ANews* | 0.048*** | 0.043*** | 0.043*** | 0.043*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $Avol_{t-1}$ | 0.271*** | 0.349*** | 0.364*** | 0.356*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| *HLtoH* | 2.790*** | 4.510*** | 4.358*** | 4.473*** |
| | (0.118) | (0.116) | (0.120) | (0.119) |
| *Spread* | -24.294*** | -77.291*** | -83.699*** | -76.839*** |
| | (2.983) | (2.943) | (3.042) | (3.002) |
| *IntraVola* | 236.119*** | 313.824*** | 309.825*** | 313.384*** |
| | (4.503) | (4.443) | (4.593) | (4.532) |
| $Ret_t$ | -0.414*** | -0.087** | 0.076* | -0.072* |
| | (0.044) | (0.043) | (0.045) | (0.044) |
| $Abs(Ret_t)$ | 6.208*** | 3.601*** | 3.920*** | 3.719*** |
| | (0.074) | (0.073) | (0.076) | (0.075) |
| $Ret_{t-1:t-5}$ | -0.315*** | -0.297*** | -0.353*** | -0.325*** |
| | (0.020) | (0.020) | (0.020) | (0.020) |
| $Ret_{t-6:t-53}$ | 0.203*** | 0.155*** | 0.132*** | 0.150*** |
| | (0.007) | (0.007) | (0.007) | (0.007) |
| $\sigma_{t-1:t-5}$ | -2.251*** | -3.505*** | -3.542*** | -3.549*** |
| | (0.083) | (0.082) | (0.085) | (0.084) |
| *52HighDum* | 0.078*** | 0.026*** | 0.032*** | 0.028*** |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| *52LowDum* | 0.094*** | 0.024*** | 0.030*** | 0.025*** |
| | (0.005) | (0.005) | (0.005) | (0.005) |
| *Time FE* | Yes | Yes | Yes | Yes |
| *Firm FE* | Yes | Yes | Yes | Yes |
| $R^2$ | 0.513 | 0.619 | 0.603 | 0.612 |
| *Observations* | 121,267 | 121,267 | 121,267 | 121,267 |

# $SOI$ – Complete regression output

| | (1) Retail | (2) LR | (3) EMO | (4) CLNV |
|---|---|---|---|---|
| *Constant* | -1.337*** | -1.140*** | 3.960*** | 2.785*** |
| | (0.055) | (0.107) | (0.089) | (0.082) |
| *ASVI* | 0.196*** | -0.704*** | 0.240*** | -0.270*** |
| | (0.039) | (0.074) | (0.062) | (0.057) |
| *ANews* | -0.151*** | -0.345*** | -0.329*** | -0.365*** |
| | (0.022) | (0.043) | (0.036) | (0.033) |
| *Savol* | 0.269*** | -0.682*** | -0.046* | -0.291*** |
| | (0.010) | (0.031) | (0.026) | (0.024) |
| *HLtoH* | 35.682*** | -108.148*** | 14.638*** | -54.417*** |
| | (3.328) | (6.377) | (5.330) | (4.894) |
| *Spread* | 726.958*** | -3,113.035*** | 2,887.498*** | -1,992.808*** |
| | (73.839) | (142.253) | (118.903) | (109.169) |
| *IntraVola* | 827.030*** | 5,223.041*** | -5,845.260*** | -1,099.409*** |
| | (118.001) | (227.592) | (190.235) | (174.661) |
| $Ret_t$ | -7.564*** | 31.578*** | -0.599 | 20.398*** |
| | (1.240) | (2.368) | (1.979) | (1.817) |
| $Abs(Ret_t)$ | -35.725*** | 15.395*** | 2.108 | 8.318*** |
| | (2.190) | (4.104) | (3.430) | (3.149) |
| $Ret_{t-1:t-5}$ | 0.359 | 10.911*** | 6.166*** | 8.417*** |
| | (0.562) | (1.078) | (0.901) | (0.827) |
| $Ret_{t-6:t-53}$ | -2.265*** | -10.141*** | -1.837*** | -6.845*** |
| | (0.195) | (0.375) | (0.313) | (0.287) |
| $\sigma_{t-1:t-5}$ | -8.423*** | -24.847*** | -21.418*** | -33.778*** |
| | (2.212) | (4.246) | (3.549) | (3.259) |
| *52HighDum* | 0.245*** | 0.086 | 0.756*** | 0.545*** |
| | (0.093) | (0.178) | (0.149) | (0.137) |
| *52LowDum* | -0.086 | 2.951*** | -1.599*** | 0.924*** |
| | (0.152) | (0.291) | (0.243) | (0.223) |
| *Time FE* | Yes | Yes | Yes | Yes |
| $R^2$ | 0.018 | 0.026 | 0.030 | 0.033 |
| *Observations* | 115,298 | 115,298 | 115,298 | 115,298 |