Similar Stocks

by*

Wei He Yuehan Wang Jianfeng Yu

March 2021

Abstract

Similarity between two stocks is measured by the distance between their characteristics such as price, size, book-to-market, return on assets, and investment-to-assets. We find that after a stock's most similar stocks have experienced high (low) returns in the past month, this focal stock tends to earn an abnormally high (low) return in the current month. The long-short portfolio strategy sorted on similar-stocks' past average return earns a monthly CAPM alpha of 1.25% and a Fama-French six-factor alpha of 0.85%. This similarity effect is robust after controlling for style investing and a wide range of well-known firm-level characteristics that can predict returns in the cross section. Our result is consistent with the increased propensity for investors to buy other stocks with similar characteristics after experiencing positive returns for a currently held stock. We also explore other potential explanations for our findings.

JEL Classification: G12

Keywords: Similarity, Style Investing, Mispricing

* We thank Zhuo Chen for helpful comments. All errors are our own. Author affiliation/contact information:

Wei He is from Southwestern University of Finance and Economics. Yuehan Wang and Jianfeng Yu are from PBCSF, Tsinghua University. Email Addresses: Wei He: Doje89@gmail.com; Yuehan Wang: wangyh.16@pbcsf.tsinghua.edu.cn; and Jianfeng Yu: yujf@pbcsf.tsinghua.edu.cn.

1 Introduction

In this paper, we propose a method to measure the similarity between two stocks by the distance between their characteristics such as price, size, book-to-market, profitability, and investment rates. The past return of a stock's similar stocks could affect the future return of this focal stock for several reasons. First, investors might believe that firms with similar characteristics should earn similar returns on average, or investors could simply use the past realized returns of similar stocks as an approximation for the future return distribution of the focal stock. Thus, investors tend to buy stocks whose similar stocks just experienced high returns in the past, thus driving up these focal stocks' future returns. Second, after observing a stock with a high recent return, investors—especially retail investors who have not previously owned this stock—might seek to buy similar stocks, whose past returns may not be as high, in order to chase the trend. This behavior would again exert higher demand for these similar stocks and lead to a higher future return for stocks with high similar-stock returns. Third, after experiencing a high return in their stock investment, investors could pay more attention to stocks with similar characteristics because investors tend to prefer stocks with certain characteristics, and different investors have different preferences.[1] Thus, the high return in the currently held stock attracts investor attention, and this attention tends to spill over to similar stocks because of investor preferences. Lastly, the psychology literature argues that personal experience should have a large impact on personal decisions (Weber et al., 1993; Hertwig et al., 2004) and suggests that when investors have positive (negative) prior investment experience in some stocks, they are more (less) likely to buy new stocks with similar characteristics. Indeed, Huang (2019) finds that after experiencing a positive return from a stock holding, investors' propensity to buy other stocks with similar characteristics such as size, book-to-market, and industry also increases. In sum, all of the above arguments imply that if a stock's similar stocks have experienced high returns in the past, total demand for this focal stock tends to increase, and thus the subsequent return should be higher as well.

This paper studies asset pricing implications based on similarity investing. In particular, we investigate the predictive power of the average past returns of a stock's similar stocks on this focal stock. We measure the similarity between two stocks by the distance between their characteristics such as firm size, book-to-market, return on assets, investment-to-assets, and price level. For example, if two stocks have similar values of these characteristics, the

[1]For example, retail investors tend to prefer low-priced stocks (e.g., Kumar and Lee, 2006), whereas institutional investors tend to prefer high-priced stocks (e.g., Gompers and Metrick, 2001).

distance between them is small, and they are similar to each other. We choose the first four characteristics since Fama and French (2015) show that a five-factor model based on these underlying characteristics can account for almost all of the existing asset pricing anomalies. We also include the price level since Green and Hwang (2009) find that investors categorize stocks based on price. We find that after similar stocks have experienced high (low) returns in the past month, the focal stock subsequently tends to earn abnormally high (low) returns. The long-short portfolio strategy sorted on similar-stocks' past one-month returns earns a monthly CAPM alpha of 1.25% and a Fama-French six-factor alpha of 0.85%. This result is also robust after controlling for a wide range of factors and firm-level characteristics.

While these results are consistent with the view that past stock returns influence investors' demand for similar stocks, thus leading to a positive relation between the average returns of similar stocks and the future returns of the focal stock, other confounding effects can also be driving this positive correlation. First, there might be a positive contemporaneous correlation between the return of the focal stock and the average returns of similar stocks. The higher average return for this focal stock could then be driven by the traditional momentum effect. To control for this momentum effect, we include the focal stock's own past 12-month return as control variables, and our results are robust to this inclusion. We also show that similar-stock average returns over the past 1 month and past 12 months can both positively predict future returns of the focal stock. However, the predictive power of the average past 12-month return of similar stocks is subsumed by the traditional momentum factor. On the other hand, the predictive power of the average past 1-month return of similar stocks is almost unaffected by the inclusion of the traditional momentum factor. Thus, to distinguish our study from earlier literature, we choose to focus on the predictive power of the average return of similar stocks over the past 1 month.

We also investigate the role of style investing in our findings. Barberis and Shleifer (2003) argue that investors tend to simplify portfolio decisions by grouping assets into styles or categories and then allocate funds at the category level rather than across individual assets. Many studies (e.g., Teo and Woo, 2004, and Froot and Teo, 2008) indeed find evidence for positive feedback trading at the style level.[2] In addition, Lewellen (2002) and Wahal and Yavuz (2013) find that there is a momentum effect at the style level, probably due to the positive feedback trading at the style level. For example, if small/growth firms perform

[2]As style investors respond to common shocks such as investor sentiment and move funds from one category to another, their coordinated demand is likely to induce excess comovements for stocks in the same category. Numerous subsequent studies (e.g., Barberis, Sheifer, and Wurgler, 2005, and Green and Hwang, 2009) find support for their model regarding excess comovement.

well in the past 6 months, they also tend to outperform in the subsequent 6 months. To rule out the possibility that our similarity-based predictability is driven by the well-known style investing effect, we perform Fama-MacBeth regressions, where we directly control for style returns as in Wahal and Yavuz (2013), and industry returns at various horizons as well. We show that the average past returns of similar stocks still have significant predictive power. This result is probably as expected since early studies tend to focus on the predictive power of medium-term style returns (i.e., 6- to 12-month style returns), whereas we focus on the predictive power of similar-stock returns over a short horizon (i.e., a monthly horizon). Moreover, we show that even if we control for 1-month past style returns, our similarity effect remains quantitatively similar.

Since we focus on the predictive power of similar-stock returns over the previous month, we also examine the relation between our similarity effect and the traditional short-term reversal effect. In particular, one might think that if a firm's similar stock just earned a high return in the past month, the focal stock itself may also have experienced a high return. Because of the short-term reversal effect, our similarity effect should be stronger after we control for the short-term reversal. Thus, we construct a new measure of similar-stocks' average returns by regressing the original measure on the focal stock's past one-month return and take the residual as our orthogonalized measure of similar-stock average returns. Based on this orthogonalized residual measure, our similarity effect indeed becomes slightly stronger: the annualized CAPM alpha spread is 1.44% (t-stat = 7.75), and the Fama-French six-factor alpha is 1.09% (t-stat = 4.89).

To further ensure that our similarity effect is not completely driven by known anomalies in the literature or by the selection of our characteristics in constructing our similarity measure, we perform a broad set of robustness checks. First, we perform both double-sorting portfolio analysis and Fama-MacBeth regressions to control for a large set of firm-level characteristics including size, book-to-market, past returns over various horizons, return on assets, investment-to-assets, idiosyncratic volatility, Amihud illiquidity, and past industry returns over various horizons. We find that our similarity effect remains significant after all of these controls. Second, we keep only firm size and book-to-market characteristics when measuring similarity. Our main results remain similarly strong. Third, we perform a placebo test by using the average return of less similar stocks as our predictor. We find that the predictive power of the past average return of less similar stocks is indeed weaker. Fourth, we also extend our analysis to 38 international countries, and in general our main results remain significant.

Like the momentum effect, the similarity effect could also be driven by either initial underreaction or continued overreaction or a combination of these two. More specifically, as argued earlier, it could be possible that high similar-stock returns induce unjustifiably high demand for the focal stock, thus leading to continued overreaction. On the other hand, similar stocks could share similar fundamentals. If the focal stock's similar stocks earned a high return in the past month, it might be because these similar stocks have favorable fundamentals. The focal stock should also have favorable news about its fundamental, and thus the focal stock should earn a high return in the future because of the initial underpricing in the portfolio formation period. Thus, we follow the standard approach by examining the long-horizon effect to tease out these alternative underlying channels. We find that the similarity effect completely reverses in the long run when the Fama-French six-factor model is used to compute alpha. However, this effect does not reverse much when we use the CAPM or the Fama-French three-factor model to calculate abnormal returns. Since we do not have a strong prior on which factor model is more appropriate in calculating abnormal returns (i.e., the classic "joint hypothesis" problem), our evidence is mixed and suggests that the similarity effect simply could be a joint result of both initial underreaction and continued overreaction. To further tease out the underlying channel, we also analyze subsequent earnings announcements returns and forecast errors for different similarity-based portfolios. We find that stocks with high past similar-stock returns tend to have higher earnings surprises and higher earnings announcement returns, consistent with the initial underreaction channel, although this evidence itself cannot rule out the continued overreaction channel.

To further inspect the underlying mechanism, we investigate the subsequent retail order imbalance behavior for portfolios with different average similar-stock returns. We find that compared with stocks with low average similar-stock returns, the subsequent retail order imbalance for stocks with high similar-stock returns increases more significantly, reflecting stronger demand for stocks with high similar-stock returns. This finding nicely complements the evidence in Huang (2019) that individuals tend to buy stocks with similar characteristics after personally experiencing high returns. While Huang (2019) finds that the same investors tend to buy stocks with similar characteristics after positive personal investment experience, our retail order imbalance includes all of the investors, not necessarily those who have already owned similar stocks before.

Lastly, we investigate the heterogeneity of our similarity effect. Since retail investors tend to have stronger beliefs about the perceived, possibly spurious, link among similar stocks, we expect that our effect is stronger among firms with low institutional holdings. Indeed, we find that the monthly CAPM alpha spread is 1.15% among firms with institutional ownership and only 0.37% among firms with high institutional ownership. The stronger similarity effect among firms with low institutional ownership is also consistent with arbitrageurs' great difficulty in shorting stocks among these firms (see, e.g., Nagel, 2005). In addition, we perform subperiod and other subsample analysis. We find that our similarity effect is significant in both early and recent samples, among both large and small stocks, and among firms with both high and low levels of idiosyncratic volatility, although the effect is stronger among smaller stocks, stocks with high idiosyncratic volatility, and in the early sample period. These results are consistent with the view that the similarity effect is stronger among firms with more retail investors and among firms that are more difficult to arbitrage.

Perhaps the study most closely related to ours is that of Huang (2019), who finds that after investors personally experience positive returns from their stock holdings, their propensity to buy other stocks in the same industry and stocks with similar size and book-to-market also increases. Huang (2019) focuses on investor trading behavior, whereas our study focuses on the asset pricing implications of similarity investing, namely, the predictive power of the average returns of the focal stock's similar stocks in the cross section. We find that similarity investing can produce large alpha for portfolio strategies based on similar-stock average returns, even after controlling for a wide range of firm-level characteristics.

Our paper is also related to the general literature on style investing and categorical thinking. Earlier studies (e.g., Barberis and Sheifer, 2003, and Barberis, Sheifer, and Wurgler, 2005) on category thinking argue that, to simplify portfolio decisions, many investors first group stocks into categories such as small-value stocks or high-tech industry stocks, and then allocate funds at the level of these categories rather than at the individual stock level. Thus, category investment can lead to excess comovement among firms in the same category. Peng and Xiong (2006) provide a theoretical foundation for categorical thinking due to attention constraints. Many subsequent studies find a supporting role for category investing in the price formation process and especially the excess comovement. For example, Green and Hwang (2009) find evidence supporting the view that investors categorize stocks based on price, and they show that stocks that undergo splits experience an increase in comovement with low-priced stocks and a decrease in comovement with high-priced stocks. Distinct from the style investing literature, however, our similarity effect does not require category investing; thus, investors could still invest at the asset level rather than at the category level. All we require is that investors are more likely to buy new stocks with similar characteristics after a stock experiences a high return in the recent past. Moveover, we show that our similarity effect remains significant after controlling for style investing.

More specifically, using mutual fund data and eight years of institutional trading data, Teo and Woo (2004) and Froot and Teo (2008), repectively, find evidence on institutional investors' style investing behavior. Different from these studies, however, we argue that our results are more likely to be driven by retail investors rather than institutional investors since the similarity effect is stronger among firms with lower institutional ownership. Relatedly, Wahal and Yavuz (2013), Lewellen (2002), and Teo and Woo (2004) find that style returns have predictive power for individual stock returns. We show that our similarity effect remains significant after controlling for various past style returns. In addition, earlier studies tend to focus on the predictive power of style returns measured over longer horizons such as 12 months. For example, Wahal and Yavuz (2013) show that style returns measured over the prior 12 months are significant predictors of future returns, but style returns measured over shorter horizons have less significant power. Our study focuses on the predictive power of the average return of similar stocks measured over the past month rather than the past 12 months. More important, Wahal and Yavuz (2013) find that the predictive power of style returns is stronger over the recent subsample, probably because there are more style investors among mutual funds and other institutional investors in the recent sample period.[3] Prior to 1988, Wahal and Yavuz (2013) find that style returns have almost no predictive power for individual stock returns. In sharp contrast to this finding, we show that our similarity effect is stronger in the early period and weaker in the recent period since retail trading shows a declining trend over time.

Our study is also related to the literature on economically linked firms. Prior studies have documented cross firm return predictability among economically related firms, including firms that are linked along the supply chain (Cohen and Frazzini, 2008; Menzly and Ozbas, 2010), single- and multi-segment firms operating in the same industries (Cohen and Lou, 2012), firms operating in the same product markets (Hoberg and Phillips, 2018), firms with similar technologies (Lee et al. 2019), and firms headquartered in the same geographic regions (Parsons, Sabbatucci, and Titman, 2020). In addition, Jin and Li (2020) find that the past returns of geography-linked firms predict focal firm returns. In this paper, we focus on simple characteristics such as size, book-to-market, profitability, and investment, which might be more easily accessible to individual investors than the previous, more sophisticated information on economic links such as the supply chain. More important, prior studies

[3] Another possibility is that demand from retail investors is driving the style investing of some mutual funds. In this sense, the retail investors are still the ones who are also partially driving the predictive power of style returns in the more recent sample period.

typically hypothesize that a firm's fundamental and stock performance should comove with its economically and geographically linked peer firms, thereby lending support to under-reaction to information regarding economically and geographically linked firms, whereas our evidence suggests potential continued overreaction to the performance of similar stocks. Moreover, our similarity effect is robust after controlling for these economic/geographical links.

Lastly, our study is also related to investor attention and positive feedback trading. Numerous studies have found a significant role for inattention and positive feedback trading in asset pricing anomalies.[4] As Barber and Odean (2008) point out, trading requires attention, and investors face a substantial search problem when choosing stocks to buy. In this paper, we hypothesize that, after a stock experiences a high return, investors tend to pay more attention to stocks with similar characteristics, probably because investors pay attention to stocks with certain characteristics. Thus, this attention leads to a higher demand and a higher future return for stocks with high average similar stock returns. The positive feedback channel also implies that investors tend to increase their positions after a positive investment experience (Pearson, Yang, and Zhang, 2021). In fact, investors tend to trade more intensively after a positive investment experience, even if the positive experience comes from winning IPO lotteries purely by chance (Ben-David, Birru, and Prokopenya, 2018; Anagol, Balasubramaniam, and Ramadorai, 2021; Gao, Shi, and Zhao, 2021). Here, different from earlier studies, we argue that after observing positive returns for a stock, investors are particularly more likely to trade stocks with firm-level characteristics that are similar to the original high-return stocks.

The rest of the paper is organized as follows. We describe data sources and variable definitions in Section 2. Section 3 presents our main results using portfolio analysis and Fama-MacBeth regressions. Section 4 further explores several alternative channels. Section 5 concludes.

2 Data Description and Summary Statistics

We describe the data sources and variable construction in this section. Our data come from several sources. Stock return data are from Center for Research in Security Prices (CRSP)

[4]This literature is too voluminous to summarize here, but readers can see Daniel and Hirshleifer (2015), Barber, Lin, and Odean (2019), and Gabaix (2019) for reviews.

from 1963 to 2019. Accounting data are from Compustat. Institutional ownership data are from the Thomson Financial 13F file from 1980 to 2019. The transaction data are from the Institute for the Study of Securities Markets (ISSM) from 1983 to 1992 and the Trade and Quote (TAQ) covering 1993 to 2000. Our stock return and accounting data for non-US companies come from the Compustat Global database. Our sample includes all NYSE, AMEX, and NASDAQ (CRSP exchange codes 1, 2, and 3) listed ordinary common stocks (CRSP share codes 10 and 11). To control for the effect of microcaps, our main analysis relies on value-weighted returns. Following Campbell et al. (2008), to alleviate the effect of a bid-ask bounce, we also delete stocks with a price less than $1 at the portfolio formation date.

2.1 SIM Definition

Our key variable is similarity between two stocks (SIM). At the end of each month t, we demean and standardize all n characteristics (price, size, book-to-market, operating profitability, and investment) so that they have a unit standard deviation. In this n-dimensional space, for each stock, we compute the Euclidean distance to all other stocks. For example, the five characteristics of stock i are (Prc_i, $SIZE_i$, BM_i, OP_i, INV_i), so the distance between stock i and stock j is

$$D_{i,j} = \sqrt{(Prc_i - Prc_j)^2 + (SIZE_i - SIZE_j)^2 + (BM_i - BM_j)^2 + (OP_i - OP_j)^2 + (INV_i - INV_j)^2} \tag{1}$$

Then, SIM is defined as the value-weighted average excess return of its 50 nearest stocks. Notice that we do not include past returns in our characteristics list since doing so would effectively bring SIM close to the focal stock's own return, and sorting on SIM would behave the same as sorting on the focal stock's own past returns.

The construction of characteristics is from standard procedure. Price is measured as the stock price at the end of each month in dollars. Size is measured as the logarithm of market capitalization (in million dollars) at the end of June in each year. Following Davis, Fama, and French (2000), we construct book equity as stockholders' book equity, plus balance sheet deferred taxes (Compustat item ITCB) and investment tax credit (TXDB) if available, minus the book value of preferred stock. The book-to-market ratio is computed as the book equity for the fiscal year ending in calendar year $t-1$ divided by the market capitalization at the end of June in year t. Following Fama and French (2015), we measure operating profit in year $t-1$ as year $t-1$ gross profit (Compustat item GP), minus selling, general, and administrative expenses (XSGA) if available, minus interest expense (XINT) if available, all divided by year $t-1$ book equity. We measure investment as the change in total assets from the fiscal year ending in year $t-2$ to the fiscal year ending in $t-1$, divided by $t-2$ total assets at the end of each June.

2.2 Control Variables

To test whether our similarity effect is driven by other known anomalies, we control for a large number of firm-level characteristics, including past 1-month return, R1m, and cumulative return from month $t-12$ to month $t-2$, R12m. Long-term reversal (Lt rev) is the cumulative return from month $t-60$ to month $t-13$. IVOL is idiosyncratic volatility as defined in Ang et al.(2006) using the daily return in the previous month. Max is the maximum daily return in month t-1, as in Bali, Cakici, and Whitelaw (2011). Skew is the skewness of monthly returns over the previous five years. Coskew is coskewness, computed as in Harvey and Siddique (2000) using five years of monthly returns.

We follow Wahal and Yavuz (2013) to construct the style return. Book-to-market and market capitalization are defined as above. We calculate book-to-market breakpoints using the full set of securities. However, to mitigate the impact of microcap firms, we use NYSE size breakpoints, which are downloaded from Kenneth French's website. These breakpoints, established at the end of each June, are used to construct 5×5 size and book-to-market style portfolios, which we then use to construct style returns (1 month, 6 months, and 1 year) for each style portfolio. In some sense, these style portfolio returns are calculated in a manner similar to the DGTW benchmark. Lastly, industry momentum is computed as the value-weighted return (1 month, 6 months, and 1 year) within the industry portfolios, which are formed with Fama-French 48-industry classifications.

2.3 Retail Order Imbalance

The measure of retail order imbalance (RIMB) is inferred from the transaction data from ISSM and TAQ. We employ the filter from Hvidkjaer (2006) and only include NYSE and AMEX common stocks from 1983 to 2000. Trades and quotes with irregular terms are excluded, and likely erroneous prices are excluded by a simple price-based error filter.

To begin with, we follow the algorithm from Lee (1992) to classify all trades by size. Specifically, at the end of month $t-1$, all stocks are grouped based on their firm size. Then, all the trades in month t are classified as small, medium, or large trade using the following dollar value breakpoints for each size-quintile group:

Firm-size quintile	Small	2	3	4	Large
Small trade cutoff, in $	3400	4800	7300	10300	16400
Large trade cutoff, in $	6800	9600	14600	20600	32800

All trades are further classified as buy-initiated or sell-initiated based on the Lee and Ready (1991) algorithm. This standard procedure uses trade price relative to the current bid-ask quotes to determine trade direction. A trade is classified as a buy-initiated (sell-initiated) if it is executed at a price above (below) the quote midpoint. If it is executed right at the quote midpoint, then we classify the trade as buy-initiated (sell-initiated) if the trade price is above (below) the last executed trade price.

According to the previous two steps, all trades are classified by size and direction. Retail order imbalance (RIMB) is then computed as the relative difference between buy-initiated and sell-initiated trade volume in a small trade group:

$$RIMB = \frac{BUYVOL - SELLVOL}{BUYVOL + SELLVOL} \tag{2}$$

For each stock, we measure the change in the retail (large) order imbalance by taking the difference between the daily imbalance and the average daily imbalance of the six-month window from month $t-7$ to month $t-2$. Then, we calculate the cumulative retail order imbalance, and to better demonstrate the results, we substract the average order imbalance from the order imbalance of five groups in each trading day (or month) to detrend. In subsequent figure 2, we present the detrened retail order imbalance.

2.4 International Sample

Following Gao, Parsons, and Shen (2018), our international sample includes common stocks traded on 43 stock exchanges from 38 countries. All return and account variables are converted to US dollar so we also employ filters at the country level. A country is included in that month or week only if at least 50 stocks have valid values for stock return. Following Hou, Karolyi, and Kho (2011), we also require a stock to have a minimum of 12 monthly observations in our sample period to be included. To alleviate the distortion of small cap firms, we exclude micro-cap firms that have a price or market capital below 5% in each month (week) in a country. We also take two steps to avoid extreme returns, following Ince and Porter (2006): 1) any return above 300% that is reversed within one month is treated as missing. Specifically, if R_t or R_{t-1} is greater than 300%, and $(1+R_t)*(1+R_{t-1})-1 < 50\%$, then both R_t and R_{t-1} are set as missing; 2) returns are set as missing if they fall out of the 0.1 and 99.9 percentiles. Our international sample starts in 1999 and ends in 2019.

3 Empirical Analysis: The Predictive Power of Similar-Stock Returns

In this section, we first perform our main portfolio analysis based on our key variable SIM. Then we use double-sorting and Fama-MacBeth regression analysis to control for other compounding effects. In addition, we also study the long-horizon return behavior after portfolio formation to tease out potential underlying mechanisms. Lastly, we investigate the subsequent earnings announcement returns and forecast errors after portfolio formation periods.

3.1 Unconditional Sorts Based on the Average Returns of Similar Stocks

At the beginning of each month, we sort stocks into decile portfolios based on each stock's 50 most similar stocks' average return in the past month or past year. We then hold the portfolio for a month and then rebalance it at the beginning of each month. Table 1 reports various characteristics of the portfolio based on similarity. In particular, we first compute the mean values of various firm-level characteristics across all stocks in the decile for each portfolio.

This table then reports the time-series average of firm-level characteristics from 1963 to 2019. The main variables are the sorting variable SIM, the average return of similar stocks in month $t+1$ (SIM_{t+1}), the stock price ($Price$), the log of market capitalization ($Size$), the market beta, the book-to-market ratio ($BMratio$), the $Profitability$, the $Investment$, idiosyncratic volatility ($IVOL$), past 1-month and 12-month stock return ($R1m, R12m$), and an illiquidity measure ($Amihud$). As we can see, firms with high SIM tend to have slightly lower beta, higher price level, larger market capitalization, higher book-to-market, higher profitability, slightly lower investment rates, lower IVOL, and lower Amihud illiquidity. Interestingly, firms with high SIM also tend to have higher past one-month and one-year returns, suggesting that firms with similar characteristics tend to have correlated returns. We also report the subsequent one-month average returns for the similar stocks for each stock in each decile portfolios (SIM_{t+1}). The average return for similar stocks is actually higher for the high-SIM portfolio, a pattern in sharp contrast to the standard short-term reversal effect.

Table 2 reports our main results on portfolios sorted on SIM. Panel A and Panel B report the portfolio spreads for equally weighted and value-weighted portfolios, respectively. The results are similar across two methods, and thus we focus on the value-weighted returns in our subsequent analysis to mitigate the impact from microcap stocks. More specifically, Panel B indicates that the stocks whose similar stocks earned high returns in the past month tend to earn higher returns than the stocks whose similar stocks earned low returns in the past month. The raw return spread between the portfolio with high similar-stock returns and the portfolio with low similar-stock returns is 1.09% per month (t-stat = 5.19). We also report the alpha spread under various factor models. All the alphas are economically large and statistically significant. For example, the Fama-French six-factor alpha is 0.85% per month (t-stat = 3.49).

In Panel C, we repeat the exercise in Panel B by replacing similar-stock returns in the past month with similar-stock returns in the past year. We also obtain large raw return spreads of 0.76% per month (t-stat = 3.13). However, after controlling for the momentum factor, most of the abnormal return is gone and the Fama-French six-factor alpha is only 0.01% per month, which is statistically insignificant (t-stat = 0.06). Thus, in our subsequent analysis, we focus on the predictive power of similar-stock returns over the past month. This approach also distinguishes our study from earlier studies on style investing, which typically focus on the predictive power of style returns over longer horizons, such as the one-year horizon in Wahal and Yavuz (2013) and Teo and Woo (2004). For example, Wahal and Yavuz (2013) show that style returns measured over the prior 12 months are significant predictors of future returns, but the style returns measured over shorter horizons have insignificant predictive power. In Panel D, we also repeat the exercise in Panel B by replacing similar-stock returns in the past month with similar-stock returns in the past week. Again, the similarity effect remains significant with an annualized return spread of 11.70% (t-stat = 5.31).

In Panel E, we also calculate similarity using only one single characteristic one by one. The results indicate that the similarity effect is still significant based on a single characteristic. However, the economic magnitude is much smaller than similarity based on multiple characteristics. For example, the raw return spread is 0.58%, 0.35%, 0.56%, 0.52%, and 0.40% per month for similarity based on price, book-to-market, size, profitability and investment, respectively. This result might be expected since the measure based on several important characteristics simultaneously might capture the similarity between stocks for general investors better than the measure based on any one specific firm characteristic. Thus, it is important to use multiple characteristics to measure similarity between stocks. We also report the results by omitting price level in calculating similarity. This way, only the four characteristics underlying the Fama-French five factors are used in computing similarity between stocks. The six-factor Fama-French alpha spread is still 0.81% per month (t-stat = 3.61). Since size and book-to-market are the two most famous firm-level characteristics, we use size and book-to-market only to calculate similarity. The portfolio alpha spread based on this similarity measure is also large at 0.92% per month (t-stat = 4.72).

Huang (2019) documents that after experiencing positive returns from holding a stock, investors' propensity to buy stocks in the same industry significantly increases. Thus, to investigate whether our similarity effect is completely driven by the trading behavior documented in Huang (2019) or by the industry momentum effect documented in Grinblatt and Moskowitz (1999), we form portfolios using the past month average return of the 50 most similar stocks in different industries relative to the focal stock. The results are reported in the last row of Panel E. The similarity effect still remains large and significant at 0.62% per month (t-stat = 2.34). Thus, our similarity effect is not completely driven by investors' higher propensity to buy stocks in the same industry after experiencing a high return from a stock investment, or the classic industry momentum effect. In our subsequent analysis, we shall control for the industry momentum effect more directly with a Fama-MacBeth regression.

In Panel F, we perform a placebo test. Instead of forming portfolios based on the most similar 50 stocks, we use the previous month's average returns of the 50 stocks with a similarity distance starting from the 51st, 101st, 151st, and 201st similar stock of the focal stock. In general, as we use a less similar-stock's average return to form portfolios, the return spread becomes smaller, as expected. If we randomly choose a focal stock's peer stocks and use the average return of these randomly selected peer stocks to form portfolios, the return spread is insignificant. These exercises suggest that the past returns of more similar stocks indeed contain more useful information about the focal firm's future return than the past returns of less similar stocks.

The existence of a short-term reversal effect is well known. Table 1 shows that stocks with high similar-stock returns also tend to have high past returns. Thus, the short-term reversal effect implies lower subsequent returns for these stocks. To purge out the traditional short-term reversal effect, in Table 3, we perform the same exercise as in Table 2, but now we orthogonalize the focal stock's similar-stock average returns against the focal stock's own past month return. As we can see, the similarity effect becomes even stronger, with a raw return spread of 1.31% per month (t-stat = 7.01) and a Fama-French six-factor alpha of 1.09% per month (t-stat = 4.89). We also report abnormal returns using alternative prominent models (e.g., factor models based on Hou, Xue, and Zhang (2015), Stambaugh and Yuan (2017), and Daniel, Hirshleifer, and Sun (2020)), and the results are still significant. To save space, in subsequent analysis, we only report Fama-French factor model adjusted returns.

3.2 Double Sorting

To further ensure that our similarity effect is not completely driven by known anomalies documented in the literature, in this subsection, we perform a series of standard double-sorting portfolio analyses to control for these existing anomalies one by one. In the next subsection, we use a Fama-MacBeth regression to control for many existing anomaly variables simultaneously.

More specifically, following the double-sorting procedure in Barberis et al. (2016), at the beginning of each month, we divide all firms in our sample into five groups based on a lagged anomaly variable such as book-to-market, and within each of the anomaly groups, we further divide firms into five portfolios based on the focal stock's similar-stock average return. The portfolio is then held for one month, and value-weighted excess returns are calculated. We then calculate the average similarity quintile portfolio return across all anomaly quintiles. It is also impossible to control for all of the anomaly variables, and it is probably even unnecessary since the Fama-French six-factor model can account for most of the existing anomalies in the literature. Thus, we specifically control for the list of characteristics underlying the Fama-French six factors.

Table 4 shows that the similarity effect remains significant after controlling for each of these anomaly variables, including prominent return predictors such as firm size, book-to-market, investment rates, profitability, IVOL, Amihud illiquidity, and past returns over various horizons and the industry momentum. Overall, our similarity effect is distinct from the existing anomalies. Notice that we use quintile portfolios rather than decile portfolios; the return spreads are smaller in magnitude, ranging from 0.40% to 0.85% per month, but still remain both economically meaningful and statistically significant. In particular, we show that our results are robust after controlling for past returns over various horizons and the industry momentum effect. Thus, our similarity effect is not completely driven by the correlation between SIM and these past-return-based variables.

3.3 Fama-MacBeth Regressions

The previous subsections report the average portfolio returns in single-way sorting and double-sorting exercises. Although the above double-sorting approach is simple and intuitive, it cannot explicitly control for other variables that might influence returns simultaneously since dependent sorting along more than three variables is infeasible because of the limited number of stocks. Therefore, we conduct an alternative analysis using predictive regressions to investigate whether the similar-stock average returns of the focal stock predicts returns in ways that are consistent with the portfolio analysis. The regression approach allows us to conveniently control for other compounding factors simultaneously (e.g., various anomaly variables and style returns in particular), which enables us to check that the similarity effect documented in the previous subsection is not just a result of comovement with other well-known anomalies.

Besides controlling for the standard anomaly variables as in our double-sorting exercise, we are particularly concerned with the potential correlation of our SIM variable with the traditional style returns. Barberis and Shleifer (2003) argue that investors tend to simplify portfolio decisions by grouping assets into styles or categories and then allocate funds at the category level rather than across individual assets. Teo and Woo (2004) and Froot and Teo (2008) indeed find evidence for positive feedback trading at the style level. This positive feedback trading could lead to the return predictive power of past style returns. Indeed, Wahal and Yavuz (2013) find that a firm's style return can predict its future returns in the cross section. For example, if a firm belongs to small/growth style, and if this style has earned high returns in the past year, this firm tends to have a higher return in the subsequent year.

Size and value are probably the most popular style investing in practice and are widely used in the investment management community (for example, by mutual funds and pension funds). In addition, Kumar (2009) and Froot and Teo (2008) find that retail and institutional investors allocate capital at the style level and do so using size and value-growth dimensions. Here, we follow Wahal and Yavuz (2013) to calculate these style returns (details are provided in Section 2).

To rule out that our predictability is driven by the well-known style investing effect documented in Wahal and Yavuz (2013), we perform Fama-MacBeth regressions, where we directly control for style returns, as defined in Wahal and Yavuz (2013), and industry returns at various horizons as defined in Grinblatt and Moskowitz (1999) as well. Table 5 reports the results of regressing excess returns on the lagged SIM variable and various control variables. First, consistent with Wahal and Yavuz (2013), column (2) indicates that style returns over the past one month and six months have insignificant predictive power, whereas the style returns over the past year have significant and positive predictive power for future stock returns. Thus, our focus on the predictive power of the average past one-month return of similar stocks also further distinguishes our study from the style investing literature.

To further tease out the effect of industry momentum from our similarity effect, in column (3) we control for industry returns over various horizons. We find that industry returns over the past 1 to 12 months have significant and positive predictive power, consistent with Grinblatt and Moskowitz (1999). In addition, our similarity effect remains significant after controlling for the industry momentum effect. In column (4), we control for the traditional predictors such as size, book-to-market, investment rate, profitability, past returns over various horizons, IVOL, and Amihud illiquidity simultaneously, and we find that SIM is still a significant predictor for future returns. Lastly, columns (5-7) show that the average past returns of similar stocks still have significant predictive power after controlling for all of these effects simultaneously. This result is likely expected since early studies tend to focus on the medium-term style momentum effect, whereas we focus on the short-term effect. Nonetheless, we show that even if we control for the past one-month style return, our similarity effect remains quantitatively similar.

In sum, we show that our similarity effect remains significant after controlling for a large set of compounding variables. In particular, we should note that our results are robust after controlling for style investing and industry momentum.

3.4 Similarity Effect over the Long Horizon

Although we motivate our tests with the natural investing behavior based on similarity noted in the introduction, this similarity effect has several potential explanations. To further distinguish alternative channels underlying our findings on the similarity effect, we examine the portfolio return spread based on similarity over longer horizons. In particular, we want to see whether reversal takes place over longer horizons.

Our similarity effect has at least two possible channels. One is continued overreaction after portfolio formation, and the other is initial underreaction at the portfolio construction date. Our results could be driven by both channels. The driving force in the overreaction channel is the additional attention that spills over to the focal stocks or the belief that firms with similar characteristics should move together—possibly because of the (mistaken) belief that similar stocks also have similar fundamentals. For example, if some stocks experienced high returns in the past, investors might pay more attention to stocks with similar characteristics and create a higher demand for these stocks. Thus, stocks with similar characteristics tend to experience a higher return in the future.

Alternatively, investors may believe that the focal stock should have a higher fundamental since its similar stocks just experienced high returns, leading to temporarily higher prices. Additionally, when investors—especially retail investors—make investment decisions, they tend to believe that similar stocks should earn similar returns. Thus, they could use the realized return of similar stocks in the past as the distribution for the stocks' future return. For example, if one finds that stock A experienced a high return in the past month while a similar stock B experienced a relatively lower return, investors might think that stock B could be undervalued and thus increase their demand for stock B, again driving up the price for stock B. All of these beliefs could be mistaken, however, and investors' subsequent higher demand for stocks with high similar-stock returns is manifested as continued overreaction after portfolio formation. Lastly, Huang (2019) finds that after personally experiencing a positive rather than negative investment return from a stock, investors' propensity to buy stocks in the same industry or stocks with similar size and book-to-market significantly increases. This could also lead to higher demand for stocks with high similar-stock returns and exert subsequent price pressure, again leading to continued overreaction. Without detailed survey data on beliefs, it is almost impossible, as well as beyond the scope of our study, to differentiate the above subchannels for continued overreaction. Instead, we simply document the asset pricing effect due to similarity investing. If this continued overreaction is the main channel underlying the similarity effect, then we should observe long-term reversal.

For the underreaction story, similar stocks might genuinely share similar fundamentals. After all, the similarity between two stocks is measured as the distance between firm characteristics, where several characteristics are firm fundamentals such as profitability and investment. Thus, if a stock's similar stocks have earned a high (low) return in the past month, it is possible that the focal stock was undervalued (overvalued) at portfolio formation, and the subsequent higher (lower) return is a price correction to the fundamental value. If this channel is completely responsible for the documented similarity effect, then there should be no subsequent long-term reversals. Thus, we can examine the long-term similarity effect to distinguish these alternative channels.

Figure 1 presents the long-term effects, and here we find mixed evidence. If particular, whether long-term reversals take place depends on the factor model used to calculate the portfolio alphas. For the Fama-French three-factor model, the reversal effect is very small. For the Fama-French six-factor model, however, there is a complete long-term reversal. Since we cannot rule out the possibility that some of the underlying factors capture mispricing rather than risk, we have no prior on which factor model is the most appropriate benchmark in adjusting for risks. Thus, our evidence on the long-term similarity effect is inconclusive. In addition, the cumulative alphas under the Fama-French six-factor model at horizons longer than three months are no longer significant, suggesting at least partial long-term reversal.

In addition, Table 6 reports the monthly average alphas of portfolios formed on SIM over various holding periods. For a holding period longer than one month, for instance (e.g., six months), the implication is that for a given decile in each month, there are six subdeciles, each of which is initiated in a different month in the prior six-month period. We take the simple average of the subdecile returns as the monthly return of the decile. As a comparison, in Panel B, we report the long-horizon results of the momentum strategy. As we can see, under all of the factor models, the average alpha for SIM-sorted portfolios is positive in all horizons. However, for momentum portfolios, the raw return spreads and CAPM alpha spreads are negative at the six-year horizon. On the other hand, under all of the other models, the average alphas of our SIM-sorted long-short portfolios are lower than those for the corresponding momentum portfolios. In addition, for horizons longer than five years, none of the SIM-sorted portfolios produce significant alphas, except for the alpha under the Fama-French three-factor model, which is only marginally significant. To the extent that momentum is partially driven by continued overreaction, the above evidence suggests that our similarity effect is at least partially driven by overreaction. In the next section, we examine subsequent earning announcement returns and analyst forecast errors to shed more light on the underlying channels.

3.5 Subsequent Earnings Announcements

To further inspect the underlying mechanism behind our findings, we examine the subsequent earnings announcement returns and earnings forecast errors. In Table 7, we perform regressions of daily firm-level stock return $Dret_{i,d}$ for firm i on day d in month t onto firm i's SIM in month $t-1$, $SIM_{i,t-1}$, the earnings announcement date dummy variable ($Eday$) on firm i on day d in month t, and their interaction term. An earnings announcement window is defined as the one-day window (columns 1 and 2) or three-day window (columns 3 and 4) centered on an earnings announcement date. The earnings announcement date dummy is one if day d in month t is in this window. Following Engelberg et al. (2018), we define the earnings announcement day as the day with the highest volume. We control for the day-fixed effect and other lagged control variables including lagged values for each of the past five days for stock returns and trading volume. We find that SIM is positively related to future daily returns, consistent with our earlier portfolio and Fama-MacBeth regression results. In addition, stock returns are higher on earnings announcement dates, consistent with the well-known earnings announcement premium. Indeed, many studies, including Ball and Kothari (1991), Cohen et al. (2007), Frazzini and Lamont (2007), and Barber et al. (2013), have documented an earnings announcement premium. Lastly, we find that the subsequent earnings return is indeed significantly larger for the portfolio with high similar-stock returns. This finding is consistent with a behavioral story in which initial expected earnings are too low (high) relative to the rational benchmark for the firms with high (low) similar-stock returns. However, it could also be consistent with a risk-based story in which more uncertainty is associated with the earnings announcement periods and thus returns are higher return during such periods.

We also examine the subsequent forecast errors in Table 8, which reports forecasting regressions of the next quarter's standardized unexpected earnings (SUE) on SIM. We include the firm fixed effect and the year-quarter fixed effect in columns (1) and (2). In column (3), we include the industry fixed effect and the year-quarter fixed effect. We also add one- to four-quarter lags of the firm's own SUEs as control variables. Table 8 shows that the earnings surprise for stocks with high similar-stock returns tends to be larger than those with low similar-stock returns, with t-statistics of 1.82 when all controls are included. This evidence is consistent with the initial underreaction interpretation, although it is only marginally significant. That is, firms with high similar-stock returns are initially undervalued because of pessimistic earnings forecasts, and the subsequent earnings announcements at least partially correct these instances of underpricing.

Although we find evidence supporting the initial underreaction channel, we cannot completely rule out the overreaction channel. In fact, as we shall show later, the retail order imbalance tends to be larger for firms with high similar-stock returns, suggesting continued overreaction of these investors. In addition, we would like to point out again that, based on the Fama-French six-factor alpha, we observe a long-term reversal, suggesting continued overreaction. Thus, our similarity effect could be driven by a combination of initial underreaction and subsequent overreaction.

3.6 Retail Order Imbalance

The evidence in the previous subsection is consistent with the initial underreaction channel. However, it does not rule out the continued overreaction channel. In this subsection, we study whether individual investors' demand for stocks with high similar-stock returns subsequently increased, since the continued overreaction channel would imply a higher demand for high-SIM stocks from retail investors.

As Barber and Odean (2000, 2008) point out, investors face a substantial search problem when choosing stocks to buy. Thus, investors might pay more attention to stocks in the same category or stocks with similar characteristics. Indeed, using data from households' detailed trading records at a large discount broker between 1991 and 1996 (Barber and Odean, 2000), Huang (2019) finds that after experiencing a positive return from their stock investments, investors are more likely to buy stocks in the same industry and stocks with similar size and book-to-market value. Thus, if a stock's similar stocks experienced a high return in the past month, it is more likely that demand for this stock from the existing holders of those similar stocks will increase, probably leading to a higher subsequent return for the focal stock. Thus, the evidence in Huang (2019) provides the microfoundation for our analysis. In some sense, our study explores the asset pricing implications of such investor trading behavior, as documented in Huang (2019).

Huang (2019) studies the demand for a stock from investors from only one discount broker, and from only the existing investors of other similar stocks. Below we investigate the trading behavior of all retail investors, including new market participants. In particular, we examine the subsequent retail order imbalance for similarity-based portfolios following the portfolio formation period. We use the algorithm from Lee and Ready (1991) to calculate the retail order imbalance, as discussed in detail in Section 2. Figure 2 shows that, indeed, the demand from retail investors increases more or decreases less for stocks with high similar-stock returns in the past month than for stocks with low similar-stock returns in the past month, consistent with and extending the evidence in Huang (2019). This higher subsequent demand for stocks with high similar-stock returns could be partially responsible for the higher subsequent returns for these stocks.

We also perform the same exercise for the large, potentially institutional, order imbalance in Figure 2. The pattern is the same as in the retail order imbalance, but the overall effect is much weaker. Again, this increased retail and large order imbalance for stocks with high similar-stock returns is consistent with both continued overreaction and initial underreaction. More specifically, the higher demand—especially from retail investors—for stocks with high similar-stock returns could be a result of investors' mistaken beliefs on the future returns of these stocks or of investors' enhanced attention to these stocks, leading to continued overreaction. On the other hand, if firms with high similar-stock returns are relatively more underpriced than firms with low similar-stock returns, then both retail and institutional imbalances are rationally responding these relative mispricing. However, since individual investors should be slower than institutional investors in responding to the initial mispricing relative to fundamental values at the portfolio formation period, the stronger subsequent demand from retail investors for high similar-stock returns suggests that this enhanced retail order imbalance for stocks with high similar-stock returns is potentially more consistent with the continued overreaction channel.

Table 9 also reports the average monthly detrended retail (large) order imbalance in the subsequent n trading months. At the end of each month t, we form 5 portfolios based on SIM. For each stock, we measure the change in the retail (large) order imbalance by taking the difference between daily order imbalance(IMB) and the average daily IMB of the six-month window from month $t-7$ to month $t-2$. Then, we calculate the cumulative retail order imbalance. To detrend the order imbalance over time, we remove the average order imbalance from the order imbalance of the five groups in each trading day. As we can see, the retail order imbalance for stocks with high SIM is significantly higher than that for stocks with high SIM for the next one to six months. For horizons up to three months, the large order imbalance for stocks with high SIM is also higher than that for stocks with low SIM. However, the difference is quite small and statistically insignificant. In addition, for the six-month horizon, the large order imbalance for stocks with high SIM is less than that for stocks with low SIM.

4 Further Robustness Checks and Alternative Channels

This section provides a few robustness checks including subperiod and subsample analysis and extends the analysis to international settings.

4.1 Subsample Analysis

In Table 10, we perform a subsample analysis by splitting the whole sample into two subsamples. It shows that our similarity effect is stronger in the early subsample. More specifically, the monthly Fama-French three-factor alpha is 1.36% in the early subsample (t-stat = 5.28) and 1.11% in the recent subsample (t-stat = 3.44). This is consistent with the notion that there is less retail trading in the more recent sample and thus our similarity effect is weaker since individuals are more subject to similarity trading. This finding is in sharp contrast to style investing. In fact, Wahal and Yavuz (2013) find that style returns have predictive power for individual stock returns in the cross-section only in the recent subsample after 1988. Prior to that, the predictive power of past style returns is mostly indistinguishable from zero. In that latter period, which coincides with increased use of size and value categorization in mutual funds and institutional portfolios, the slopes on style returns are large and reliably positive.

We would like to point out that the ultimate investors in these mutual funds are still mostly retail investors. Thus, one could still argue that the availability of style investing means that retail investors' style investing affects asset prices, especially in the more recent sample period. Even if this is the case, it would still imply that the increased retail participation in style mutual funds amplifies the predictive power of the style return in the recent sample period, whereas the decreased direct stock market participation of retail investors reduces the predictive power of similar-stock returns in the more recent sample period. Thus, similarity investing is distinct from the style investing, although two concepts are related.

In addition, in Table 10, we split our sample into two subsamples by firm size. The market cap in the group of big firms is above the median NYSE size, whereas that in the group of small firms is below the median. We find a significant similarity effect across both groups of firms. Wahal and Yavuz (2013), however, do not find any predictability of past style returns among big stocks (those above the median NYSE size), implying that style returns based on value growth alone do not help explain the cross-sectional variation in returns among large stocks. In particular, we find that the similarity effect among the big group of firms is smaller but remains statistically significant with a monthly alpha of 0.68% (t-stat = 4.08).

Lastly, we also split our sample into two equal subsamples by IVOL. We find a significant similarity effect across both groups of firms. Not surprisingly, the similarity effect is smaller among firms with low IVOL but is still statistically significant with an alpha of 0.69% per month (t-stat = 4.73). Thus, our similarity effect is not limited to small firms and firms that are difficult to arbitrage, although the effect is indeed weaker among these firms, consistent with the idea that limits to arbitrage prevent the correction of mispricing. However, arbitrageurs may not be fully aware of this similarity investing effect, and thus they do not trade against this effect enough to eliminate the observed similarity effect among the low-IVOL or large-size group.

4.2 Interaction with Institutional Ownership

If retail investors are more subject to similarity investing, then we expect that the similarilty effect should be more pronounced among firms with lower institutional ownership. Table 11 reports the average benchmark-adjusted returns for portfolios constructed by sequentially sorting on institutional ownership and SIM. The high-IO (low-IO) subsample consists of the top (bottom) half of stocks sorted on size-adjusted IO, computed by following Nagel (2005).

The CAPM alpha of the similarity-based strategy is 1.15% per month among firms with low institutional ownership, whereas the CAPM alpha of the similarity-based strategy is only

0.37% per month among firms with high institutional ownership. The difference of 0.78% per month is also statistically significant (t-stat = 4.31). A similar, albeit slightly weaker, pattern also emerges for the Fama-French three-factor alpha and the Fama-French six-factor alpha. Using mutual fund data and eight years of institutional trading data, Teo and Woo (2004) and Froot and Teo (2008), respectively, find evidence on institutional investors' style investing behavior. Different from the predictive power of style returns, which is more likely to be driven by institutional investors,[5] we argue that our similarity effect is more likely to be driven by retail investors since the similarity effect is stronger among firms with lower institutional ownership.

4.3 Evidence from International Markets

In this subsection, we examine the international data to see whether the results we documented are an international phenomenon. Since different countries have different levels of turnover, they also have a different volume-based clock. For example, for countries with high turnover, one week would correspond to one month relative to countries with low turnover. For countries with extremely high turnover, such as China, investors may pay more attention to the past week's return rather than the past month's return. Thus, we divide the countries into two groups by their aggregate turnover, which is the average annual turnover across all the years for which we have data for that country. For turnover higher than the cross-sectional median, we use the past one-week average return of similar stocks as our predictive variable, whereas for turnover below the median, we still use monthly returns of similar stocks as our predictor. In addition, we use aggregate turnover as a proxy for individual investor ownership in each country since individual investors tend to trade more. The country-level aggregate turnover is the average annual turnover across all the years for which we have data for the country.

We repeat the exercise for the 38 countries in Table 12. We find that among the countries in the high turnover group, 17 out of 19 have a positive return spread based on similarity, and 10 of them are statistically significant. Our international sample starts in 1999 and ends in 2019, so it is much shorter than our US counterpart, and thus the statistical significance for individual countries is typically weaker, as expected. However, the economic magnitude for the average portfolio is still large, with an annualized return spread of 12.17% (t-stat

[5]Ultimately, it is individual investors who invest in style mutual funds. Thus, they may also be the ultimate forces behind the predictive power of style returns. Nonetheless, mutual fund managers also have discretion over stock selections.

= 7.91). In addition, only 2 of these 19 countries have a negative return spread, but none of them is statistically significant. On the other hand, the average return spread for the group with low turnover is indeed smaller and statistically and economically insignificant, with an annualized spread of 1.92% (t-stat = 1.22), consistent with the notion that higher participation from retail investors tends to amplify the similarity effect.

5 Conclusions

In this study, we propose a measure of similarity between two stocks based on the distance between their corresponding characteristics. We find that firms with high similar-stock returns tend to earn significantly higher future returns than firms with low similar-stock returns. We argue that this finding is consistent with the view that investors' demand for firms with high similar-stock returns increases because of similarity-based investing behavior. We also show that our similarity effect is not completely driven by style investing and other well-known anomalies. Our evidence indicates that our similarity effect could be consistent with both initial underreaction and continued overreaction.

References

Anagol, Santosh, Vimal Balasubramaniam, and Tarun Ramadorai, 2021, Learning from noise: Evidence from India's IPO lotteries, *Journal of Financial Economics*.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *The Journal of Finance* 61.1, 259–299.

Bali, Turan G., Nusret Cakici, and Robert F. Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99.2, 427–446.

Ball, Ray, and Stephen P. Kothari, 1991, Security returns around earnings announcements, *Accounting Review* 718–738.

Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *The journal of Finance* 55.2, 773–806.

Barber, Brad M., and Terrance Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *The review of financial studies* 21, 785–818.

Barber, Brad M., Emmanuel T.De George, Reuven Lehavy, and Brett Trueman, 2013, The earnings announcement premium around the globe *Journal of Financial Economics* 108, 118–138.

Barber, Brad M., Shengle Lin, and Terrance Odean, 2019, Mediating Investor Attention, Working paper, University of California, Berkeley.

Barberis, Nicholas, and Andrei Shleifer, 2003, Style investing, *Journal of financial Economics* 68, 161–199.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283–317.

Barberis, Nicholas, Abhiroop Mukherjee, and Baolian Wang, 2016, Prospect theory and stock returns: An empirical test, *The Review of Financial Studies* 29.11, 3068–3107.

Ben-David, Itzhak, Justin Birru, and Viktor Prokopenya, 2018, Uninformative feedback and risk taking: Evidence from retail forex trading, *Review of Finance* 22, 2009–2036.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of political Economy* 107, 205–251.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *The*

Journal of Finance 63.6, 2899–2939.

Chen, Huafeng, Shaojun (Jenny) Chen, Zhuo Chen, Feng Li, 2019, Empirical investigation of an equity pairs trading strategy, *Management Science* 65, 370–389.

Cohen, Daniel A., Aiyesha Dey, Thomas Z.Lys, Shyam V.Sunder, 2007, Earnings announcement premia and the limits to arbitrage, *Journal of Accounting and Economics* 43.2, 153–180.

Cohen, Lauren, and Andrea Frazzini, 2008, Economic links and predictable returns, *The Journal of Finance* 63, 1977–2011.

Cohen, Lauren, and Dong Lou, 2012, Complicated firms, *Journal of financial economics* 104, 383–400.

Daniel, Kent, and David Hirshleifer, 2015, Overconfident investors, predictable returns, and excessive trading, *Journal of Economic Perspectives* 29, 61–88.

Daniel, Kent, David Hirshleifer, and Lin Sun, 2020, Short-and Long-Horizon Behavioral Factors, *Review of Financial Studies* 33.4, 1673–1736.

Davis, James L., Eugene F. Fama, and Kenneth R. French, 2000, Characteristics, covariances, and average returns: 1929 to 1997, *The Journal of Finance* 55.1, 389–406.

Engelberg, Joseph, R. David McLean, and Jeffrey Pontiff, 2018, Anomalies and news, *The Journal of Finance* 73.5, 1971–2001.

Fama, Eugene, and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Frazzini, Andrea, and Owen A. Lamont, 2007, The earnings announcement premium and trading volume, NBER working paper w13090.

Froot, Kenneth, and Melvyn Teo, 2008, Style investing and institutional investors, *Journal of Financial and Quantitative Analysis* 883–906.

Gabaix, Xavier, 2019, Behavioral inattention, *Handbook of Behavioral Economics: Applications and Foundations* 1, 261–343.

Gao, Huasheng, Donghui Shi, and Bin Zhao, 2021, Does good luck make people overconfident? Evidence from a natural experiment in the stock market, *Journal of Corporate Finance* 101933.

Gompers, Paul A., and Andrew Metrick, 2001, Institutional investors and equity prices, *The quarterly journal of Economics* 116, 229–259.

Goyal, Amit and Welch, Ivo, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Gao, Pengjie, Christopher A. Parsons, and Jianfeng Shen, 2018, Global relation between financial distress and equity returns, *The Review of Financial Studies* 31, 239–277.

Green, T. Clifton, and Byoung-Hyoun Hwang, 2009, Price-based return comovement, *Journal of Financial Economics* 93, 37–50.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *The Journal of finance* 55.3, 1263-1295.

Hertwig, Ralph, Greg Barron, Elke U. Weber, and Ido Erev, 2004, Decisions from Experience and the Effect of Rare Events in Risky Choice, *Psychological Science* 15, 534–539.

Hoberg, Gerard, and Gordon M. Phillips, 2018, Text-based industry momentum, *Journal of Financial and Quantitative Analysis*53, 2355–2388.

Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho, 2011, What factors drive global stock returns?, *The Review of Financial Studies* 24, 2527–2574.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: an investment approach, *Review of Financial Studies* 28, 650–705.

Huang, Xing, 2019, Mark Twain's Cat: Investment experience, categorical thinking, and stock selection, *Journal of Financial Economics* 131, 404–432.

Hvidkjaer, Soeren, 2006, A trade-based analysis of momentum, *The Review of Financial Studies* 19.2, 457–491.

Ince, Ozgur S., and R. Burt Porter, 2006, Individual equity return data from Thomson Datastream: Handle with care!, *Journal of Financial* 29, 463–479.

Jin, Zuben and Frank Weikai Li, 2020, Geographic Links and Predictable Returns, Working paper, Singapore Management University

Kumar, Alok, 2009, Dynamic style preferences of individual investors and stock returns, *Journal of Financial and Quantitative Analysis* 607–640.

Kumar, Alok, and Charles MC Lee, 2006, Retail investor sentiment and return comovements, *The Journal of Finance* 61, 2451–2486.

Lee, Charles MC, 1992, Earnings news and small traders: An intraday analysis, *Journal of Accounting and Economics* 15, 265–302.

Lee, Charles MC, and Mark J. Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 46, 733–746.

Lee, Charles MC, Stephen Teng Sun, Rongfei Wang, and Ran Zhang, 2019, Technological links and predictable returns, *Journal of Financial Economics* 132, 76–96.

Lewellen, Jonathan, 2002, Momentum and autocorrelation in stock returns, *The Review of Financial Studies* 15, 533–564.

Menzly, Lior, and Oguzhan Ozbas, 2010, Market segmentation and cross-predictability of returns, *The Journal of Finance* 65, 1555–1580.

Moskowitz, Tobias J., and Mark Grinblatt, 1999, Do industries explain momentum?, *The Journal of finance* 54.4, 1249–1290.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of financial economics* 78, 277–309.

Parsons, Christopher A., Riccardo Sabbatucci, and Sheridan Titman, 2020, Geographic lead-lag effects, *The Review of Financial Studies* 33, 4721–4770.

Pearson, Neil D., Zhishu Yang, and Qi Zhang, 2021, The Chinese warrants bubble: Evidence from brokerage account records, *The Review of Financial Studies* 34, 264–312.

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563-602.

Stambaugh, Robert F., and Yu Yuan, 2017, Mispricing factors, *The Review of Financial Studies* 30, 1270–1315.

Teo, Melvyn, and Sung-Jun Woo, 2004, Style effects in the cross-section of stock returns, *Journal of Financial Economics* 74, 367–398.

Wachter, Jessica A, 2006, A consumption-based model of the term structure of interest rates, *Journal of Financial economics* 79, 365–399.

Wahal, Sunil, and M. Deniz Yavuz, 2013, Style investing, comovement and return predictability, *Journal of Financial Economics* 107, 136–154.

Weber, Elke U., Ulf Böckenholt, Denis J. Hilton, and Brian Wallace, 1993, Determinants of diagnostic hypothesis generation: Effects of information, base rates, and experience, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19, 1151–1164.

Figure 1: Long-horizon effects

This figure plots the 72-month long-short portfolio cumulative excess returns and model adjusted alphas. The sample period is from 1963 to 2019.

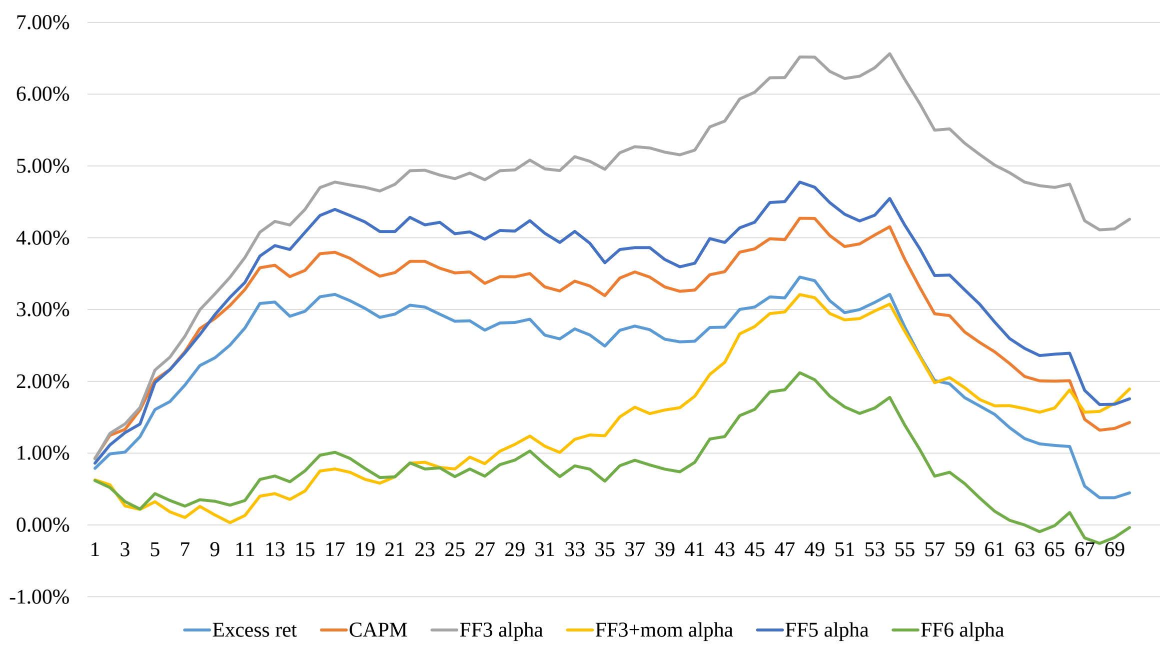

Figure 2: Subsequent retail order imbalance

This figure plots the detrended cumulative retail order imbalance in the subsequent trading month. At the end of each month t, we form five portfolios based on SIM. For each stock, we measure the change in the retail (large) order imbalance by taking the difference between the daily imbalance in month $t+1$ and the average daily imbalance of the six-month window from month $t-7$ to month $t-2$. Then we calculate the cumulative retail order imbalance. To better demonstrate the results, we substract the average order imbalance from the order imbalance of five groups in each trading day to detrend. For example, $Retail-low$ represents the average cumulative retail order imbalance for the bottom SIM quintile portfolio. The sample period is from 1983 to 2000.

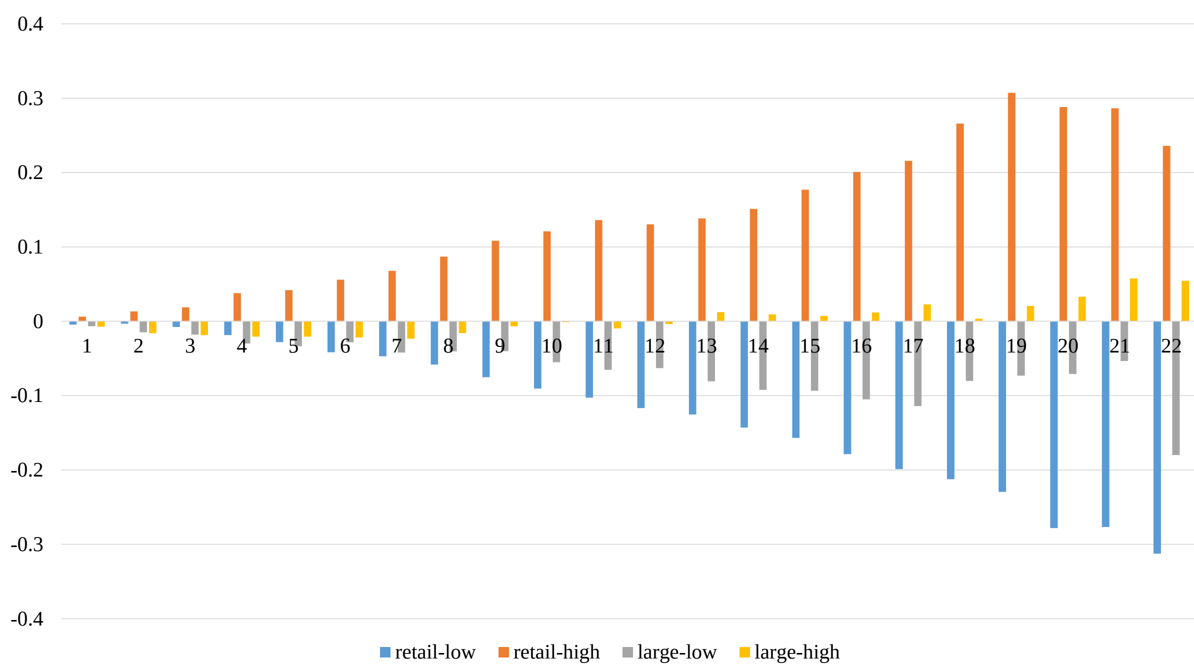

Table 1: Summary statistics

Each month, we sort stocks into deciles based on SIM and compute the mean values of the characteristics listed in the top row of the table across all stocks in the decile for each portfolio. This table reports the time-series average of firm-level characteristics from 1963 to 2019. All variables are winsorized at a 1% level. Then main variables are the sorting variable SIM, average return of matched stocks in month $t+1$ (SIM_{t+1}), the stock price ($Price$), the log of market capitalization ($Size$), the market beta, the book-to-market ratio ($BMratio$), the $Profitability$, $Investment$, idiosyncratic volatility ($IVOL$), past 1-month and 12-month stock return ($R1m, R12m$), and an illiquidity measure ($Amihud$).

	SIM	SIM_{t+1}	Market beta	Price	Size	BM ratio	Profitability	Investment	R1m	R12m	IVOL	Amihud
High	6.180	1.178	1.110	25.469	11.545	0.950	0.110	0.153	5.322	22.964	0.032	5.520
9	3.755	1.040	1.104	25.866	11.940	0.874	0.192	0.150	3.781	18.749	0.029	3.870
8	2.542	0.973	1.114	24.870	12.046	0.845	0.209	0.148	2.852	16.336	0.028	3.394
7	1.609	0.925	1.123	23.524	12.072	0.834	0.216	0.150	1.987	14.616	0.028	3.226
6	0.779	0.892	1.136	22.253	12.052	0.826	0.215	0.153	1.365	12.772	0.028	3.401
5	-0.022	0.860	1.159	20.949	12.016	0.813	0.210	0.156	0.680	10.928	0.029	3.591
4	-0.857	0.813	1.178	19.675	11.988	0.800	0.203	0.166	-0.069	9.326	0.030	3.795
3	-1.792	0.766	1.206	18.226	11.924	0.781	0.187	0.177	-0.790	7.208	0.031	4.161
2	-2.980	0.701	1.239	16.333	11.835	0.755	0.156	0.203	-1.795	5.041	0.032	4.414
Low	-5.180	0.608	1.296	13.144	11.618	0.742	0.063	0.292	-3.087	1.929	0.036	5.124

Table 2: Main results

This table reports the monthly alphas on equally weighted and value-weighted decile portfolios of stocks sorted on SIM. Specifically, for each stock at the end of each month t, we calculate the Euclidean distance to all other stocks based on standardized Price/BM/Size/Profitability and Investment. SIM is the value-weighted average excess (to the risk-free rate) return of the past $1 month$ of the nearest 50 stocks. Panel A reports the next month's alphas and the long-short strategy performance of equally weighted portfolios. Panel B reports the value-weighted results. In Panel C, SIM is calculated based on the value-weighted average excess return of the past *year*. In Panel D, SIM is based on excess return of the past *week*. Panel E reports the Fama-French six-factor adjusted alphas on the value-weighted decile portfolios. Here, for each stock at the end of each month t, we calculate the Euclidean distance to all other stocks based on one variable (Price/Size/BM/Profitability/Investment), four characteristics (except for Price), and two characteristics (BM and Size). Also reported are the results of similar stocks within and outside of the same Fama-French 48 industries. Panel E reports the placebo test with similarity distance starting from the 51st, 101st, 151st, and 201st similar stock of the focal stock. We also report the results of randomly selected peer stocks. The sample period is from 1963 to 2019. Newey-West three-lag adjusted t-statistics are reported in parentheses.

Panel A	Equal weighted										
	Low	2	3	4	5	6	7	8	9	High	High-Low
Excess return	0.152	0.447	0.530	0.700	0.796	0.838	0.901	1.009	1.092	1.303	1.151
	(-0.50)	(1.69)	(2.10)	(2.82)	(3.32)	(3.55)	(3.91)	(4.27)	(4.52)	(4.88)	(6.13)
CAPM alpha	-0.583	-0.228	-0.119	0.061	0.187	0.241	0.316	0.424	0.527	0.731	1.314
	(-3.49)	(-1.70)	(-0.95)	(0.53)	(1.61)	(2.20)	(2.94)	(3.73)	(4.04)	(4.57)	(7.19)
FF3 alpha	-0.667	-0.355	-0.265	-0.087	0.035	0.083	0.174	0.270	0.376	0.585	1.252
	(-5.34)	(-3.99)	(-3.55)	(-1.43)	(0.62)	(1.63)	(3.48)	(4.52)	(4.58)	(5.18)	(6.66)
FF3+Mom alpha	-0.34	-0.124	-0.093	0.041	0.151	0.181	0.217	0.315	0.396	0.608	0.948
	(-2.39)	(-1.20)	(-1.18)	(0.56)	(2.42)	(3.13)	(4.17)	(4.92)	(4.72)	(4.96)	(4.40)
FF5 Alpha	-0.517	-0.293	-0.248	-0.056	0.047	0.094	0.174	0.280	0.386	0.609	1.126
	(-3.19)	(-2.45)	(-2.70)	(-0.68)	(0.69)	(1.68)	(3.49)	(4.55)	(4.38)	(4.94)	(4.59)
FF6 Alpha	-0.252	-0.101	-0.102	0.051	0.144	0.177	0.212	0.318	0.404	0.628	0.881
	(-1.68)	(-0.88)	(-1.20)	(0.59)	(2.16)	(2.93)	(4.08)	(4.83)	(4.52)	(4.89)	(3.72)

Panel B	Value weighted										
	Low	2	3	4	5	6	7	8	9	High	High-Low
Excess return	-0.003	0.378	0.481	0.554	0.694	0.763	0.759	0.852	0.976	1.088	1.090
	(-0.01)	(1.60)	(2.26)	(2.72)	(3.45)	(3.86)	(3.73)	(4.21)	(4.39)	(4.58)	(5.19)
CAPM Alpha	-0.722	-0.273	-0.128	-0.044	0.110	0.202	0.194	0.296	0.419	0.529	1.252
	(-4.8)	(-2.21)	(-1.43)	(-0.54)	(1.42)	(2.78)	(2.21)	(3.41)	(3.56)	(3.92)	(6.07)
FF3 Alpha	-0.719	-0.291	-0.179	-0.071	0.074	0.190	0.219	0.282	0.413	0.505	1.224
	(-5.41)	(-2.42)	(-2.09)	(-0.95)	(1.05)	(2.88)	(2.55)	(3.32)	(3.36)	(4.05)	(5.79)
FF3+Mom Alpha	-0.497	-0.119	-0.080	0.008	0.093	0.147	0.144	0.214	0.329	0.398	0.895
	(-3.34)	(-0.75)	(-0.93)	(0.09)	(1.27)	(2.20)	(1.96)	(2.85)	(3.20)	(3.37)	(3.95)
FF5 Alpha	-0.537	-0.206	-0.208	-0.041	0.082	0.227	0.327	0.334	0.483	0.582	1.118
	(-3.61)	(-1.31)	(-2.14)	(-0.46)	(1.16)	(3.36)	(2.92)	(3.22)	(2.79)	(3.60)	(4.05)
FF6 Alpha	-0.365	-0.068	-0.120	0.023	0.097	0.187	0.251	0.271	0.406	0.484	0.849
	(-2.55)	(-0.38)	(-1.33)	(0.22)	(1.33)	(2.79)	(2.76)	(3.02)	(2.97)	(3.54)	(3.49)

Table 2 (cont.): Main results

Panel C — Yearly return based SIM

	Low	2	3	4	5	6	7	8	9	High	High-Low
Excess return	0.282	0.405	0.519	0.438	0.624	0.679	0.710	0.801	0.908	1.046	0.764
	(0.97)	(1.66)	(2.33)	(2.17)	(3.24)	(3.46)	(3.78)	(4.07)	(4.40)	(4.25)	(3.13)
CAPM Alpha	-0.432	-0.246	-0.101	-0.138	0.067	0.109	0.159	0.252	0.356	0.445	0.876
	(-2.59)	(-1.97)	(-0.92)	(-1.43)	(0.83)	(1.38)	(2.01)	(2.83)	(3.32)	(3.12)	(3.64)
FF3 Alpha	-0.491	-0.295	-0.140	-0.190	0.010	0.069	0.127	0.214	0.336	0.472	0.963
	(-3.06)	(-2.36)	(-1.35)	(-1.91)	(0.13)	(0.94)	(1.72)	(2.79)	(3.58)	(3.91)	(4.00)
FF3+Mom Alpha	0.160	0.172	0.163	-0.039	0.080	0.032	0.021	0.045	0.069	0.124	-0.037
	(1.34)	(1.44)	(1.70)	(-0.42)	(0.97)	(0.41)	(0.26)	(0.56)	(0.82)	(1.15)	(-0.21)
FF5 Alpha	-0.333	-0.241	-0.155	-0.235	-0.075	-0.021	0.032	0.091	0.310	0.522	0.855
	(-1.78)	(-1.55)	(-1.31)	(-2.17)	(-0.87)	(-0.27)	(0.42)	(1.09)	(2.72)	(3.92)	(2.96)
FF6 Alpha	0.209	0.153	0.106	-0.101	-0.006	-0.043	-0.049	-0.040	0.085	0.219	0.011
	(1.87)	(1.25)	(1.09)	(-1.05)	(-0.07)	(-0.54)	(-0.62)	(-0.49)	(0.97)	(2.07)	(0.06)

Panel D — Weekly return based SIM

	Low	2	3	4	5	6	7	8	9	High	High-Low
Excess return	-0.072	-0.038	0.040	0.027	0.040	0.086	0.087	0.123	0.118	0.158	0.230
	(-1.34)	(-0.81)	(0.89)	(0.63)	(0.90)	(1.95)	(1.94)	(2.69)	(2.44)	(2.85)	(5.31)
CAPM Alpha	-0.202	-0.162	-0.082	-0.091	-0.078	-0.031	-0.031	0.006	-0.001	0.034	0.236
	(-7.10)	(-7.97)	(-4.61)	(-5.48)	(-4.70)	(-1.91)	(-1.73)	(0.31)	(-0.03)	(1.06)	(5.37)
FF3 Alpha	-0.213	-0.173	-0.094	-0.102	-0.087	-0.041	-0.039	-0.004	-0.012	0.017	0.230
	(-7.84)	(-8.45)	(-5.26)	(-6.18)	(-5.30)	(-2.60)	(-2.44)	(-0.24)	(-0.58)	(0.63)	(5.41)
FF3+Mom Alpha	-0.187	-0.157	-0.085	-0.097	-0.085	-0.036	-0.039	-0.006	-0.010	0.028	0.215
	(-6.50)	(-7.55)	(-4.67)	(-5.65)	(-4.94)	(-2.24)	(-2.27)	(-0.33)	(-0.48)	(1.05)	(4.80)
FF5 Alpha	-0.195	-0.173	-0.097	-0.109	-0.093	-0.046	-0.038	-0.007	-0.002	0.054	0.250
	(-7.50)	(-8.82)	(-5.52)	(-6.74)	(-5.59)	(-2.86)	(-2.39)	(-0.35)	(-0.09)	(2.09)	(5.95)
FF6 Alpha	-0.172	-0.158	-0.089	-0.104	-0.091	-0.041	-0.038	-0.008	-0.001	0.063	0.235
	(-6.29)	(-7.85)	(-4.92)	(-6.21)	(-5.24)	(-2.50)	(-2.28)	(-0.41)	(-0.03)	(2.39)	(5.41)

Table 2 (cont.): Main results

Panel E	Single/multi-variable distance									
	Low	2	3	4	5	6	7	8	9	High	High-Low
Price	-0.155	-0.133	-0.229	-0.134	0.019	0.162	0.113	0.075	0.286	0.427	0.582
	(-0.73)	(-1.14)	(-1.98)	(-1.56)	(0.22)	(1.45)	(1.69)	(0.97)	(4.09)	(4.59)	(2.27)
BM	-0.105	-0.106	-0.073	-0.049	0.136	0.046	0.005	0.104	0.088	0.242	0.347
	(-1.35)	(-1.53)	(-1.14)	(-0.89)	(1.90)	(0.79)	(0.07)	(1.48)	(1.26)	(3.38)	(3.03)
Size	-0.185	0.012	-0.149	0.071	0.032	0.096	0.186	0.209	0.224	0.371	0.556
	(-1.76)	(0.13)	(-1.90)	(0.97)	(0.49)	(1.23)	(2.12)	(2.64)	(2.69)	(3.51)	(3.14)
Profitability	-0.185	-0.029	-0.001	0.001	-0.007	0.055	0.093	0.056	0.139	0.330	0.515
	(-1.92)	(-0.33)	(-0.02)	(0.01)	(-0.13)	(0.85)	(1.48)	(0.92)	(2.19)	(3.31)	(3.38)
Investment	-0.156	-0.126	0.018	-0.004	0.069	0.093	0.103	0.105	0.046	0.242	0.398
	(-1.94)	(-2.00)	(0.29)	(-0.06)	(0.98)	(1.47)	(1.48)	(1.48)	(0.66)	(3.12)	(3.35)
Without price	-0.256	-0.161	-0.193	0.111	-0.007	0.194	0.245	0.279	0.422	0.558	0.814
	(-1.92)	(-1.61)	(-2.29)	(1.48)	(-0.10)	(2.89)	(2.82)	(3.11)	(3.63)	(4.24)	(3.61)
BM/Size	-0.400	-0.319	-0.107	0.126	0.215	0.214	0.245	0.303	0.352	0.516	0.916
	(-3.44)	(-3.45)	(-1.17)	(1.60)	(3.06)	(2.84)	(3.02)	(3.17)	(3.41)	(4.71)	(4.72)
Same industry	-0.516	-0.187	-0.243	0.065	0.048	0.087	0.175	0.250	0.378	0.378	0.895
	(-4.04)	(-1.73)	(-2.39)	(0.57)	(0.56)	(1.04)	(2.21)	(2.76)	(4.02)	(3.04)	(4.30)
Diff industry	-0.166	0.038	0.005	0.080	0.096	0.226	0.110	0.268	0.273	0.457	0.623
	(-1.00)	(0.24)	(0.06)	(1.03)	(1.29)	(3.14)	(1.34)	(3.02)	(2.34)	(3.52)	(2.34)

Panel F	Robustness check									
	Low	2	3	4	5	6	7	8	9	High	High-Low
Near 51-100	-0.355	-0.226	-0.044	0.032	0.178	0.155	0.301	0.269	0.251	0.470	0.825
	(-2.60)	(-2.18)	(-0.46)	(0.36)	(2.51)	(2.20)	(3.71)	(2.50)	(2.55)	(3.57)	(3.52)
Near 101-150	-0.395	-0.091	-0.047	0.112	0.044	0.113	0.205	0.272	0.208	0.526	0.920
	(-2.94)	(-0.82)	(-0.54)	(1.33)	(0.56)	(1.76)	(2.33)	(2.97)	(2.10)	(4.49)	(4.46)
Near 151-200	-0.242	-0.062	-0.021	0.080	-0.067	0.170	0.150	0.207	0.325	0.338	0.580
	(-1.88)	(-0.54)	(-0.27)	(0.95)	(-0.91)	(2.44)	(2.21)	(2.93)	(3.65)	(2.54)	(2.62)
Near 201-250	-0.145	-0.164	-0.182	0.110	0.169	0.127	0.104	0.034	0.178	0.264	0.408
	(-1.36)	(-1.79)	(-2.00)	(1.51)	(2.23)	(1.61)	(1.40)	(0.43)	(1.82)	(2.29)	(2.27)
Random matched	0.051	0.079	0.015	0.016	0.010	0.080	0.147	-0.103	0.053	0.041	-0.010
	(0.83)	(1.27)	(0.21)	(0.23)	(0.15)	(1.36)	(2.08)	(-1.79)	(0.81)	(0.70)	(-0.11)

Table 3: Orthogonal SIM

This table reports the monthly alphas on value-weighted portfolios of stocks sorted on orthogonalized SIM. Specifically, for each stock i at the end of each month t, we regress the focal stock's similar-stock average returns against the focal stock's own past month's return and here shows the return alphas of portfolios sorted on a regression residual. Also reported are HXZ Q4, SY4, and DHS3 model-adjusted alphas. The sample period is from 1963 to 2019. Newey-West three-lag adjusted t-statistics are reported in parentheses.

	Low	2	3	4	5	6	7	8	9	High	High-Low
Excess return	-0.103	0.214	0.328	0.633	0.620	0.800	0.676	0.830	1.019	1.209	1.311
	(-0.4)	(0.93)	(1.55)	(3.20)	(3.12)	(4.28)	(3.47)	(4.00)	(4.80)	(5.15)	(7.01)
CAPM Alpha	-0.799	-0.424	-0.277	0.061	0.046	0.248	0.134	0.268	0.466	0.638	1.437
	(-5.96)	(-3.99)	(-3.33)	(0.77)	(0.56)	(3.72)	(1.90)	(2.81)	(4.64)	(4.98)	(7.75)
FF3 Alpha	-0.794	-0.432	-0.305	0.035	0.013	0.223	0.133	0.279	0.454	0.611	1.405
	(-6.6)	(-4.55)	(-3.80)	(0.48)	(0.18)	(3.38)	(2.04)	(2.97)	(4.54)	(5.31)	(7.29)
FF3+Mom Alpha	-0.580	-0.357	-0.228	0.077	0.064	0.195	0.086	0.217	0.377	0.546	1.127
	(-4.22)	(-3.35)	(-2.92)	(0.96)	(0.90)	(3.00)	(1.29)	(2.92)	(3.73)	(4.93)	(5.44)
FF5 Alpha	-0.655	-0.366	-0.315	0.024	0.012	0.237	0.178	0.354	0.508	0.666	1.321
	(-4.55)	(-3.67)	(-3.43)	(0.32)	(0.17)	(3.50)	(2.40)	(2.98)	(3.77)	(4.88)	(5.40)
FF6 Alpha	-0.487	-0.309	-0.248	0.061	0.057	0.211	0.133	0.295	0.438	0.605	1.092
	(-3.45)	(-2.96)	(-2.95)	(0.75)	(0.78)	(3.15)	(1.85)	(3.08)	(3.64)	(4.96)	(4.89)
HXZ Q4 Alpha	-0.446	-0.296	-0.259	0.081	0.043	0.271	0.183	0.400	0.540	0.758	1.188
	(-2.33)	(-2.49)	(-2.41)	(0.91)	(0.51)	(3.42)	(2.23)	(3.17)	(3.56)	(4.82)	(4.17)
SY4 Alpha	-0.414	-0.273	-0.232	0.083	0.077	0.210	0.049	0.236	0.350	0.596	1.010
	(-2.61)	(-2.33)	(-2.56)	(0.84)	(0.87)	(2.86)	(0.66)	(2.74)	(3.16)	(5.09)	(4.53)
DHS3 Alpha	-0.033	0.085	0.061	0.309	0.178	0.300	0.121	0.281	0.407	0.670	0.667
	(-0.16)	(0.57)	(0.50)	(2.61)	(1.74)	(3.09)	(1.30)	(2.30)	(2.50)	(3.82)	(2.32)

Table 4: Double sorting

Each month, stocks are sorted into quintiles based on a control variable (one of Size, BM, Profitability, Investment, Ret 1m, Ret 12m, Lt rev, IVOL, MAX, Skewness, Coskewness, and Industry R6m). Size is market capitalization. BM is the book-to-market ratio. Profitability is operating profitability, as in Fama and French (2015). R1m is the past 1-month return. R12m is the cumulative return from month $t-12$ to month $t-2$. Lt rev is the cumulative return from month t-60 to month $t-13$. IVOL is idiosyncratic volatility. Max is the maximum daily return in month $t-1$, as in Bali, Cakici, and Whitelaw (2011). Skew is the skewness of monthly returns over the previous five years. Coskew is coskewness, computed as in Harvey and Siddique (2000) using five years of monthly returns. Within each quintile, stocks are then further sorted into quintiles based on SIM. The returns of the five SIM portfolios over the next month are averaged across the five control variable quintiles. We report the FF6 alphas, on a value-weighted basis, of the five SIM portfolios and of the low-SIM minus high-SIM long-short portfolio. The sample period runs from 1963 to 2019. Newey-West three-lag adjusted t-statistics are reported in parentheses.

	Size	BM	Profitability	Investment	R1m	R12m	Lt rev	IVOL	MAX	Skewness	Coskewness	Industry R6m
High	0.405	0.350	0.333	0.318	0.539	-0.119	0.271	0.193	0.344	0.292	0.382	0.220
	(3.64)	(2.56)	(2.95)	(2.27)	(4.85)	(-1.12)	(2.19)	(2.80)	(4.22)	(3.20)	(3.22)	(2.57)
4	0.201	0.168	0.218	0.166	0.272	-0.102	0.163	0.111	0.215	0.187	0.231	0.059
	(3.13)	(1.92)	(2.53)	(1.76)	(3.86)	(-1.27)	(1.95)	(2.18)	(3.88)	(2.75)	(2.96)	(0.92)
3	0.029	0.046	0.118	0.033	0.166	-0.164	0.035	0.078	0.108	0.053	0.089	-0.012
	(0.51)	(0.95)	(1.91)	(0.65)	(2.72)	(-1.93)	(0.59)	(1.57)	(2.26)	(1.17)	(1.68)	(-0.19)
2	-0.099	-0.172	-0.110	-0.153	-0.075	-0.339	-0.109	-0.093	-0.113	-0.133	-0.142	-0.217
	(-1.42)	(-2.36)	(-1.39)	(-2.02)	(-0.99)	(-3.41)	(-1.40)	(-1.61)	(-1.94)	(-2.24)	(-2.23)	(-2.95)
Low	-0.384	-0.363	-0.464	-0.254	-0.310	-0.515	-0.259	-0.308	-0.201	-0.242	-0.290	-0.411
	(-3.60)	(-3.14)	(-4.18)	(-2.09)	(-2.66)	(-3.94)	(-2.17)	(-3.38)	(-1.69)	(-2.47)	(-2.74)	(-3.33)
High-Low	0.790	0.713	0.798	0.572	0.849	0.397	0.530	0.501	0.546	0.534	0.672	0.631
	(4.27)	(3.12)	(4.05)	(2.40)	(4.38)	(2.19)	(2.53)	(3.66)	(3.02)	(3.22)	(3.34)	(3.62)

Table 5: Fama-MacBeth Regressions

This table reports the results of the second stage of Fama-MacBeth regressions. In Model (1), only the lagged SIM is included. In Model (2), we take the style return from Wahal and Yavuz (2013) into consideration. In Model (3), we control for the industry return from Moskowitz and Grinblatt (1999). In Models (4)-(6), firm-level characteristics are added, including SIZE, BM, Profitability, Investment, IVOL, past 1-month and past 12-month stock return (R1m, R12m), and an illiquidity measure(Amihud). The sample period is from 1963 to 2019. The last two rows in the table report the average observations and average R^2 for all cross-sectional regressions. Newey-West three-lag adjusted t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SIM	0.331	0.238	0.283	0.139	0.121	0.115	0.101
	(5.41)	(4.56)	(5.02)	(4.42)	(4.08)	(3.80)	(3.53)
Style R1m		-0.003			0.029		0.012
		(-0.10)			(1.28)		(0.52)
Style R6m		0.006			-0.017		-0.023
		(0.11)			(-0.42)		(-0.59)
Style R1y		0.278			0.096		0.095
		(5.00)			(2.32)		(2.41)
Industry R1m			0.236			0.290	0.288
			(6.53)			(8.55)	(8.54)
Industry R6m			0.105			0.107	0.110
			(1.76)			(2.04)	(2.14)
Industry R1y			0.153			0.096	0.094
			(2.55)			(1.88)	(1.88)
Size				-0.209	-0.243	-0.193	-0.227
				(-4.41)	(-4.80)	(-4.15)	(-4.55)
BM				0.114	0.049	0.129	0.063
				(3.64)	(1.68)	(4.38)	(2.24)
profitability				0.077	0.078	0.077	0.077
				(3.82)	(3.86)	(3.86)	(3.89)
Investment				-0.163	-0.155	-0.168	-0.161
				(-7.44)	(-7.25)	(-7.76)	(-7.59)
R1m				-0.591	-0.596	-0.652	-0.656
				(-12.75)	(-12.90)	(-14.46)	(-14.56)
R12m				0.225	0.223	0.188	0.187
				(4.29)	(4.31)	(3.82)	(3.83)
IVOL				-0.258	-0.242	-0.270	-0.255
				(-3.33)	(-3.18)	(-3.66)	(-3.50)
Amihud				0.046	0.041	0.047	0.042
				(1.23)	(1.12)	(1.26)	(1.16)
Avg. Obs.	2088	2088	2088	2081	2081	2081	2081
R^2 Avg.(%)	1.113	2.217	2.531	5.846	6.149	6.801	7.073

Table 6: Long-Horizon results

This table shows the monthly average alphas of portfolios formed on SIM. A holding period that is longer than one month (e.g., six months), means that for a given decile in each month there are six subdeciles, each of which is initiated in a different month in the prior six-month period. We take the simple average of the subdecile returns as the monthly return of the decile. As a comparison, in Panel B, we report the long-horizon results of the momentum strategy. The whole sample is from 1963 to 2019.

Panel A	SIM						
	1 month	3 months	6 months	1 year	3 years	5 years	6 years
Excess return	0.787	0.339	0.287	0.257	0.084	0.039	0.020
	(4.29)	(2.65)	(2.98)	(3.41)	(1.71)	(1.00)	(0.57)
CAPM Alpha	0.927	0.443	0.358	0.294	0.094	0.043	0.023
	(4.97)	(3.50)	(3.76)	(3.81)	(1.88)	(1.07)	(0.64)
FF3 Alpha	0.921	0.468	0.386	0.334	0.140	0.084	0.061
	(4.63)	(3.30)	(3.81)	(4.28)	(2.91)	(2.33)	(1.84)
FF3+Mom Alpha	0.628	0.089	0.031	0.034	0.042	0.029	0.031
	(2.85)	(0.66)	(0.32)	(0.50)	(0.89)	(0.82)	(0.94)
FF5 Alpha	0.860	0.428	0.358	0.307	0.105	0.052	0.027
	(3.10)	(2.10)	(2.69)	(3.29)	(2.19)	(1.43)	(0.84)
FF6 Alpha	0.618	0.109	0.057	0.053	0.023	0.007	0.005
	(2.44)	(0.67)	(0.57)	(0.79)	(0.50)	(0.21)	(0.15)

Panel B	Momentum						
	1 month	3 months	6 months	1 year	3 years	5 years	6 years
Excess return	1.253	1.132	0.940	0.578	0.124	-0.005	-0.018
	(4.65)	(4.50)	(4.07)	(2.91)	(1.03)	(-0.06)	(-0.22)
CAPM Alpha	1.407	1.249	1.035	0.635	0.142	0.001	-0.011
	(5.70)	(5.32)	(4.70)	(3.22)	(1.18)	(0.01)	(-0.13)
FF3 Alpha	1.630	1.486	1.314	0.951	0.388	0.200	0.170
	(6.74)	(6.45)	(6.32)	(5.60)	(4.07)	(3.02)	(3.01)
FF3+Mom Alpha	0.339	0.289	0.300	0.256	0.149	0.064	0.078
	(3.13)	(3.07)	(2.57)	(2.04)	(1.65)	(0.99)	(1.37)
FF5 Alpha	1.374	1.284	1.167	0.864	0.333	0.159	0.137
	(4.40)	(4.49)	(4.69)	(4.71)	(3.59)	(2.53)	(2.56)
FF6 Alpha	0.295	0.280	0.313	0.276	0.132	0.046	0.060
	(2.62)	(2.89)	(2.56)	(2.07)	(1.49)	(0.78)	(1.15)

Table 7: Returns on Earnings Announcement Days

This table reports regressions of daily firm-level stock return $Dret_{i,d}$ for firm i on day d in month t onto firm i's SIM in month $t-1$, $SIM_{i,t-1}$, the earnings announcement date dummy variable (Eday) on firm i on day d in month t, and their interaction term. An earnings announcement window is defined as the one-day window (columns 1 and 2) or three-day window (columns 3 and 4) centered on an earnings announcement date. Eday is a dummy variable that equals one if day d in month t is in this window. Following Engelberg et al. (2018), we define the earnings announcement day as the day with the highest volume. We control for the day-fixed effect and other lagged control variables including lagged values for each of the past five days for stock returns and trading volume. The sample is from 1973 to 2019.

	(1)	(2)	(3)	(4)
	One-day window		Three-day window	
	Dret(%)	Dret(%)	Dret(%)	Dret(%)
SIM	0.326	0.445	0.327	0.443
	(18.27)	(25.10)	(18.21)	(24.84)
SIM*1(Eday)	1.312	1.393	0.433	0.527
	(16.05)	(17.16)	(8.90)	(10.91)
1(Eday)	0.207	0.225	0.104	0.120
	(37.95)	(41.42)	(31.76)	(37.02)
Lagged Controls	No	Yes	No	Yes
Day FE	Yes	Yes	Yes	Yes
R^2	0.062	0.062	0.075	0.075
Obs.	33463463	33463463	33445521	33445521

Table 8: Future Earnings Surprise

This table reports forecasting regressions of the next quarter's standardized unexpected earnings (SUE) on SIM. We include the firm fixed effect and the year-quarter fixed effect in columns (1) and (2). In column (3), we include industry fixed effect and the year-quarter fixed effect. We add one- to four-quarter lags of the firm's own SUEs as control variables. The sample period is from 1984 to 2019.

	(1) SUE_t	(2) SUE_t	(3) SUE_t
SIM_{t-1}	0.03556	0.03764	0.04368
	(2.21)	(1.68)	(1.82)
SUE_{t-1}		0.03807	0.06385
		(1.74)	(1.58)
SUE_{t-2}		0.00676	0.01910
		(0.53)	(1.12)
SUE_{t-3}		-0.00817	0.00500
		(-0.90)	(0.74)
SUE_{t-4}		-0.07389	-0.00001
		(-1.60)	(-0.92)
Constant	-0.00122	-0.0029	-0.00285
	(-58.49)	(-18.97)	(-23.71)
Firm FE	YES	YES	NO
Industry FE	NO	NO	YES
Year-quarter FE	YES	YES	YES
Obs.	208,079	124,570	122,042
Ajusted R^2	0.0737	0.0570	0.0201

Table 9: Long-Horizon Order Imbalance

This table reports the average monthly detrended retail (large) order imbalance in the subsequent n trading months. At the end of each month t, we form five portfolios based on SIM. For each stock, we measure the change in the retail (large) order imbalance by taking the difference between the daily IMB and the average daily IMB of the six-month window from month $t-7$ to month $t-2$. Then, we calculate the cumulative retail order imbalance. To better demonstrate the results, we substract the average order imbalance from the order imbalance of the five groups in each trading day to detrend. Here, 3 months means that for a given quintile in each month, there exist three subquintiles, and we take the simple average of the subquintile retail (large) order imbalance as the monthly retail (large) order imbalance. The sample period is from 1983 to 2000.

Panel A	Retail order imbalance			
	1 month	2 months	3 months	6 months
High	0.204	0.209	0.199	0.241
	(3.7)	(4.42)	(4.64)	(6.10)
4	0.094	0.107	0.120	0.162
	(3.17)	(4.23)	(4.95)	(7.15)
3	-0.016	-0.019	-0.009	-0.019
	(-0.53)	(-0.71)	(-0.42)	(-1.30)
2	-0.134	-0.137	-0.138	-0.185
	(-2.67)	(-3.25)	(-3.89)	(-6.39)
Low	-0.232	-0.254	-0.260	-0.331
	(-3.51)	(-4.35)	(-4.97)	(-7.53)
High-Low	0.436	0.463	0.459	0.572
	(4.22)	(5.26)	(5.83)	(8.68)

Panel B	Large order imbalance			
	1 month	2 months	3 months	6 months
High	0.008	-0.005	-0.013	-0.040
	(0.27)	(-0.18)	(-0.54)	(-1.77)
4	0.051	0.029	0.021	0.022
	(2.13)	(1.35)	(1.00)	(1.16)
3	0.000	0.001	-0.003	-0.016
	(0.01)	(0.07)	(-0.18)	(-1.24)
2	-0.049	-0.042	-0.026	-0.019
	(-1.7)	(-1.77)	(-1.36)	(-1.06)
Low	-0.087	-0.046	-0.034	-0.019
	(-2.20)	(-1.39)	(-1.08)	(-0.63)
High-Low	0.095	0.041	0.021	-0.020
	(1.60)	(0.81)	(0.45)	(-0.44)

Table 10: Subsample Analysis

This table reports the monthly alphas on value-weighted decile portfolios sorted on SIM. In Panel A, we divide the whole sample into two subperiods (1963-1990 and 1990-2019). Panel B and Panel C report the results of subsamples divided by the median marketsize breakpoint of NYSE or IVOL. Newey-West three-lag adjusted t-statistics are reported in parentheses.

Panel A — Two subperiods

		Low	2	3	4	5	6	7	8	9	High	High-Low
1963-1990	CAPM Alpha	-0.691	-0.358	-0.106	0.106	0.232	0.369	0.278	0.337	0.585	0.734	1.425
		(-3.82)	(-2.35)	(-0.91)	(0.99)	(2.12)	(3.74)	(2.75)	(3.09)	(4.61)	(4.34)	(5.54)
	FF3 Alpha	-0.757	-0.437	-0.183	-0.011	0.134	0.256	0.195	0.216	0.491	0.600	1.357
		(-4.43)	(-3.18)	(-1.56)	(-0.10)	(1.28)	(2.80)	(2.07)	(2.26)	(4.33)	(4.54)	(5.28)
	FF6 Alpha	-0.374	-0.158	0.021	0.025	0.152	0.231	0.130	0.227	0.421	0.536	0.911
		(-1.94)	(-0.98)	(0.15)	(0.20)	(1.29)	(2.36)	(1.28)	(2.15)	(2.81)	(3.35)	(2.92)
1990-2019	CAPM Alpha	-0.777	-0.212	-0.153	-0.189	0.001	0.045	0.115	0.253	0.264	0.343	1.120
		(-3.30)	(-1.12)	(-1.15)	(-1.60)	(0.01)	(0.43)	(0.82)	(1.93)	(1.40)	(1.69)	(3.56)
	FF3 Alpha	-0.734	-0.193	-0.173	-0.174	-0.001	0.072	0.175	0.284	0.300	0.379	1.113
		(-3.74)	(-1.07)	(-1.37)	(-1.67)	(-0.01)	(0.77)	(1.29)	(2.23)	(1.55)	(1.95)	(3.44)
	FF6 Alpha	-0.361	-0.010	-0.219	-0.126	0.010	0.045	0.254	0.251	0.327	0.406	0.767
		(-1.72)	(-0.04)	(-1.76)	(-0.86)	(0.10)	(0.48)	(1.71)	(1.78)	(1.43)	(1.83)	(2.05)

Panel B — Large/small subsamples

		Low	2	3	4	5	6	7	8	9	High	High-Low
Size<NYSE p50	CAPM Alpha	-0.646	-0.157	-0.037	0.004	0.131	0.271	0.390	0.408	0.554	0.675	1.321
		(-3.62)	(-1.07)	(-0.27)	(0.03)	(1.05)	(2.19)	(3.06)	(3.04)	(3.73)	(4.12)	(6.84)
	FF3 Alpha	-0.736	-0.289	-0.180	-0.159	-0.018	0.110	0.240	0.267	0.423	0.536	1.272
		(-5.51)	(-3.20)	(-2.18)	(-2.12)	(-0.26)	(1.61)	(3.49)	(3.16)	(4.21)	(4.45)	(6.20)
	FF6 Alpha	-0.339	-0.023	-0.011	-0.025	0.119	0.192	0.340	0.321	0.454	0.583	0.922
		(-2.09)	(-0.24)	(-0.11)	(-0.23)	(1.33)	(2.33)	(4.46)	(3.67)	(4.33)	(4.50)	(3.77)
Size>NYSE p50	CAPM Alpha	-0.383	-0.330	-0.131	-0.056	0.092	0.149	0.091	0.181	0.242	0.294	0.677
		(-3.51)	(-4.31)	(-1.78)	(-0.98)	(1.55)	(2.43)	(1.32)	(2.65)	(3.05)	(3.19)	(4.20)
	FF3 Alpha	-0.36	-0.312	-0.112	-0.060	0.085	0.128	0.070	0.193	0.263	0.318	0.678
		(-3.25)	(-3.84)	(-1.53)	(-1.02)	(1.43)	(2.00)	(1.00)	(2.80)	(3.37)	(3.53)	(4.08)
	FF6 Alpha	-0.144	-0.280	-0.070	-0.091	0.064	0.064	0.044	0.129	0.178	0.239	0.383
		(-1.29)	(-3.09)	(-0.86)	(-1.37)	(1.05)	(0.99)	(0.60)	(1.59)	(2.23)	(2.48)	(2.24)

Panel C — High/Low IVOL subsamples

		Low	2	3	4	5	6	7	8	9	High	High-Low
High IVOL	CAPM Alpha	-0.795	-0.714	-0.276	-0.418	-0.179	-0.066	-0.078	0.087	-0.009	0.226	1.021
		(-3.73)	(-3.64)	(-1.32)	(-2.10)	(-0.95)	(-0.39)	(-0.43)	(0.43)	(-0.04)	(1.08)	(4.36)
	FF3 Alpha	-0.784	-0.710	-0.288	-0.409	-0.190	-0.104	-0.084	0.098	0.046	0.199	0.983
		(-4.36)	(-4.62)	(-1.64)	(-2.51)	(-1.23)	(-0.80)	(-0.58)	(0.57)	(0.30)	(1.11)	(3.99)
	FF6 Alpha	-0.153	-0.193	0.100	-0.028	0.171	0.134	0.208	0.336	0.306	0.336	0.489
		(-0.73)	(-1.30)	(0.62)	(-0.17)	(1.06)	(0.98)	(1.65)	(2.05)	(1.97)	(1.74)	(1.53)
Low IVOL	CAPM Alpha	-0.162	-0.058	-0.023	0.011	0.102	0.175	0.287	0.288	0.282	0.519	0.681
		(-1.67)	(-0.64)	(-0.35)	(0.15)	(1.49)	(2.60)	(4.52)	(3.79)	(3.34)	(5.74)	(4.70)
	FF3 Alpha	-0.197	-0.110	-0.071	0.007	0.068	0.156	0.261	0.269	0.274	0.488	0.686
		(-2.14)	(-1.26)	(-1.10)	(0.09)	(0.96)	(2.36)	(4.00)	(3.96)	(3.32)	(5.63)	(4.73)
	FF6 Alpha	-0.076	-0.121	-0.100	-0.011	0.034	0.130	0.197	0.201	0.159	0.369	0.445
		(-0.80)	(-1.25)	(-1.38)	(-0.14)	(0.49)	(1.97)	(2.67)	(3.01)	(1.80)	(4.09)	(2.95)

Table 11: Institutional Ownership

This table reports the average benchmark-adjusted returns for portfolios constructed by sequentially sorting on institutional ownership (IO) and SIM. The high-IO (low-IO) subsample consists of the top (bottom) half of stocks sorted on size-adjusted IO, computed following Nagel (2005): each quarter, we regress the logit of the IO percentage on log(size) and the square of log(size) and take the regression residual as the size-adjusted IO. The sample period is from April 1980 to December 2019. Newey-West three-lag adjusted t-statistics are reported in parentheses.

Panel A	CAPM Alpha			Panel B	FF3 Alpha			Panel C	FF6 Alpha		
SIM	Low-IO	High-IO	High-Low	SIM	Low-IO	High-IO	High-Low	SIM	Low-IO	High-IO	High-Low
High	0.602	0.039	-0.563	High	0.693	0.099	-0.594	High	0.786	0.070	-0.716
	(3.45)	(0.33)	(-3.85)		(3.81)	(0.80)	(-4.05)		(3.93)	(0.52)	(-4.74)
4	0.474	0.083	-0.391	4	0.570	0.098	-0.472	4	0.677	0.021	-0.657
	(2.75)	(1.10)	(-2.16)		(3.17)	(1.34)	(-2.48)		(3.74)	(0.27)	(-3.52)
3	0.394	-0.049	-0.443	3	0.449	-0.022	-0.472	3	0.555	-0.038	-0.593
	(2.74)	(-0.65)	(-2.98)		(3.29)	(-0.31)	(-3.11)		(3.47)	(-0.52)	(-3.40)
2	-0.010	-0.170	-0.159	2	0.059	-0.171	-0.230	2	0.266	-0.141	-0.407
	(-0.08)	(-1.77)	(-1.14)		(0.49)	(-1.77)	(-1.54)		(1.50)	(-1.33)	(-1.91)
Low	-0.549	-0.327	0.222	Low	-0.453	-0.296	0.158	Low	-0.012	-0.159	-0.146
	(-2.83)	(-2.44)	(1.32)		(-2.70)	(-2.24)	(0.98)		(-0.06)	(-1.17)	(-0.77)
High-Low	1.150	0.366	-0.784	High-Low	1.146	0.394	-0.752	High-Low	0.798	0.229	-0.569
	(4.85)	(1.75)	(-4.31)		(4.39)	(1.76)	(-4.09)		(2.43)	(0.97)	(-2.33)

Table 12: International evidence

This table shows the excess return of portfolios formed on SIM in 38 countries. We divide countries into two groups by their aggregate turnover, which is the average annual turnover across all the years for which we have data for the country. For turnover higher than the cross-sectional median level, we use the past one-week average return of similar stocks as our predictive variable SIM, whereas for turnover below the median, we still use the monthly returns. Also reported are returns on a strategy that equally combines the strategies available within a given week or month(*Comb*).

	High turnover				Low turnover		
Country	Low	High	High-Low	Country	Low	High	High-Low
AUS	0.018	0.377	**0.36**	ARG	2.212	1.638	-0.645
	(0.16)	(3.57)	(4.57)		(2.82)	(1.85)	(-0.85)
BRA	-0.014	1.341	**1.355**	AUT	0.974	0.807	-0.163
	(-0.08)	(5.51)	(5.25)		(2.04)	(1.52)	(-0.31)
CAN	0.104	0.262	0.158	BEL	0.655	0.778	0.214
	(0.88)	(1.88)	(1.29)		(1.57)	(1.69)	(0.50)
CHE	0.142	0.346	**0.204**	CHL	1.178	0.664	-0.514
	(1.69)	(4.33)	(2.65)		(2.75)	(1.61)	(-1.62)
CHN	0.038	0.466	**0.427**	DNK	0.705	0.924	0.219
	(0.29)	(3.61)	(4.87)		(1.55)	(2.15)	(0.51)
DEU	0.088	0.189	0.101	ESP	0.023	0.734	**0.823**
	(1.10)	(2.05)	(1.01)		(0.05)	(1.69)	(2.11)
FIN	0.219	0.400	**0.181**	FRA	0.893	0.908	0.016
	(2.05)	(3.89)	(1.77)		(1.92)	(2.27)	(0.04)
GBR	0.095	0.055	-0.041	IDN	1.066	2.265	**1.199**
	(1.01)	(0.62)	(-0.49)		(1.65)	(3.07)	(1.91)
GRC	-0.003	0.069	0.072	ISR	1.106	1.263	-0.048
	(-0.02)	(0.51)	(0.62)		(2.15)	(2.79)	(-0.10)
HKG	-0.002	0.239	**0.24**	MEX	0.765	1.096	0.331
	(-0.01)	(2.36)	(2.33)		(1.63)	(2.03)	(0.69)
IND	0.196	0.431	**0.235**	MYS	0.612	0.802	0.246
	(1.36)	(3.34)	(2.03)		(1.43)	(1.97)	(0.77)
ITA	0.256	0.425	**0.169**	NLD	0.413	0.95	0.565
	(2.54)	(4.00)	(1.75)		(0.82)	(1.66)	(1.00)
JPN	0.182	0.103	-0.079	NOR	0.495	0.791	0.296
	(2.25)	(1.29)	(-1.45)		(0.87)	(1.53)	(0.65)
KOR	-0.153	0.127	**0.280**	NZL	0.978	1.976	1.036
	(-0.96)	(0.89)	(1.95)		(0.95)	(3.15)	(0.95)
SGP	0.100	0.378	0.168	PAK	1.319	1.439	0.120
	(0.95)	(2.50)	(1.54)		(2.29)	(2.40)	(0.24)
SWE	0.132	0.302	**0.170**	PHL	0.857	0.706	-0.206
	(1.16)	(2.82)	(2.06)		(1.59)	(1.03)	(-0.31)
THA	0.242	0.346	0.105	POL	0.612	0.973	0.360
	(2.27)	(3.33)	(1.19)		(1.09)	(1.72)	(0.81)
TUR	0.147	0.281	0.134	PRT	1.004	0.342	0.473
	(0.84)	(1.54)	(1.13)		(1.11)	(0.34)	(0.61)
ZAF	0.289	0.385	0.096	RUS	2.272	1.775	-1.018
	(2.37)	(3.18)	(0.93)		(1.97)	(1.81)	(-0.93)
Comb	0.109	0.348	**0.234**	Comb	0.916	1.097	0.160
	(1.57)	(4.96)	(7.91)		(2.76)	(3.41)	(1.22)