Slow-Moving Liquidity Provision and Flow-Driven Common Factors in Stock Returns

January 24, 2019

Jiacui Li
Stanford University
Graduate School of Business

Abstract

We demonstrate that retail investor demand shocks explain close to 30% of Fama-French size and value factor return variation. From 1965 to 2015, retail investors frequently made large capital reallocations across mutual funds with different factor exposures, imposing significant factor-level demand shocks on the stock market. These retail flows generated large price movements that reverse in the subsequent years. We provide evidence that the flows generate large price movements due to slow-moving liquidity provision: other investors are slow to learn about and trade against the flow-induced dislocations. Existing heterogeneous agent models with frictionless liquidity provision cannot explain the observed patterns.

Keywords: Slow-Moving Capital, Common Factors, Mutual Fund Flows, Demand Effects

JEL classification: G10, G12, G23, G40

I am grateful for comments from Yu An, Nick Barberis, Simcha Barkai, Jonathan Berk, Svetlana Bryzgalova, John Cochrane, Darrell Duffie, David Hirshleifer, Lunyang Huang, Yan Ji, Arvind Krishnamurthy, Ken Li, Hanno Lustig, Tim McQuade, Markus Pelger, Cameron Peng, Matthew Ringgenberg (discussant), Charles Shi (discussant), Ken Singleton, Yang Song, Ashish Tiwari (discussant), Edward Watts, Yilin Yang, Darya Yuferova (discussant), and Eric Zitzewitz. I also thank participants at WFA 2018, SFS Cavalcade 2018, AFA 2018 (poster session), AFBC 2017, and AFM 2017. All errors are mine. Please address correspondence to: Jiacui Li, Stanford Graduate School of Business, 655 Knight Way, Stanford, CA 94305. E-mail: jiacui@stanford.edu
1 Introduction

A central question in asset pricing is understanding what drives stock price movements. Fama and French (1996) demonstrate that two style factors, size and value, explain a large percentage of stock price movements. However, the drivers of those two factors are not well understood. In particular, Fama and French (1995) find that factor movements are only weakly linked to factor level cash flow innovations. In other words, there is an “excess volatility puzzle” in these factors.

In this paper, we show that retail investors capital reallocation at the style level can explain close to 30% of the size and value factor movements. From 1965 to 2015, retail investors frequently reallocated hundreds of billions of dollars across mutual funds with different investment styles (e.g. from small-cap funds to large-cap funds). Because fund managers trade stocks within their designated styles in response to investor flows (Lou (2012)), these mutual fund flows directly lead to large style-level stock market order imbalances. These imbalances cause temporary price pressures: when retail investors buy or sell a style, the corresponding stock factor appreciates or depreciates contemporaneously, followed by price reversions in the subsequent years (Figure 1).

Figure 1. Mutual fund flows forecast Fama-French factor return reversals. Horizontal axis plots the annual net flow into each style factor and vertical axis plots the annualized factor return in the subsequent five years. Each data point is an annual observation from 1965 to 2015. For both factors, flows negatively predict subsequent factor returns with statistical significance (section 2.5). Green dashed lines are the best linear fits to the data. The computation of flows into each factor is explained in section 2.4.

To begin with, we document that mutual fund flows have strong factor structures in size and value. Following the methodology in Fama and French (1996), we double sort funds by their size and value factor loadings into five by five portfolios. Principal component analysis reveals that flows into these 25 portfolios have a strong factor structure. After removing a “market” component, the
two largest principal components appear to be linear combinations of the size and value factors, and they explain 26% and 18% variance, respectively. Moreover, the flow factor structure is not explained away by the fact that fund flows chase past returns.

These style-level mutual fund flows are large enough to create price pressures in the stock market. As an example of the magnitude, from 1997 to 1999, retail investors moved $332 billion from small-cap funds to large-cap funds, a reallocation of 2.4% of total U.S. equity market value. During 2001 to 2005, investors reallocated $640 billion from growth funds to value funds, which is 4.1% of the total market value. Because fund managers respond to investor purchases and redemptions by expanding or reducing holdings (Lou (2012)), the style-level fund flows directly lead to large style-level trading in the underlying stocks.

To analyze the price impact of flows at the style level, we use mutual fund flows to construct size (small-minus-big) and value (high-minus-low) flow factors in the same way Fama and French construct return factors. These flow factors measure the net flow into each of the return factors. Consistent with flows creating temporary price dislocations, flow factor movements are associated with large contemporaneous return factor movements and subsequent price reversions. For both factors, one standard deviation higher (or lower) flow leads to around 3% lower (or higher) annual factor returns in the subsequent five years. The reversion predictability is both statistically and economically significant with \( R^2 \) in the 17-27% range.

Why can fund flows generate such large price movements? We provide evidence that liquidity provision is slow-moving: some of the other investors, such as pension funds, are slow to learn about and trade against the flow-induced price dislocations. When flow shocks occur, only a subset of liquidity providers, such as broker-dealers, are aware of the flows and can immediately provide liquidity. Therefore, some of the initial price movements are excessive and will revert over time as the other investors gradually learn about the flow-induced investment opportunity. We formally test for slow-moving liquidity under the guidance of a simple model. The key distinguishing feature of slow-moving liquidity is that flows will forecast price reversions even if flows do not mean-revert. This is different from existing “rational” and “behavioral” models with frictionless liquidity provision which generate price reversions only through mean-reverting flows\(^2\). The empirical results are consistent with slow-moving liquidity.

To quantify the flow-induced price effects, we fit structural vector autoregressive models to the joint dynamics of factor flows and returns. For both factors, we find that flows can explain almost 30% of factor return variation. This high explanatory power should not come as a surprise as the reversion forecasting regressions already generate \( R^2 \)s around 20%, and the total effect on returns also includes the initial price pressures. Comparing our estimates to the existing literature, we find that the estimated price impact coefficients in our data are similar to those in well-identified

\(^2\) Most models with heterogeneous agents assume frictionless liquidity provision, and the “rational” or “behavioral” designation is only about the nature of the liquidity demand. For instance, in “rational” models such as Chan and Kogan (2002) and Dumas (1989), the trading demands arise from shocks to wealth, preferences, or income. In “behavioral” models such as Barberis and Shleifer (2003) and Barberis, Greenwood, Jin, and Shleifer (2015), demands come from irrational investor beliefs.
index inclusion studies. The speed of price reversion (one to three years) is also consistent with other papers on temporary price impacts of mutual fund flows (Coval and Stafford (2007), Edmans, Goldstein, and Jiang (2012)).

Our work is closely aligned to a few papers showing that demand/supply pressures can create broad market-level price swings. In the treasury bond market, Krishnamurthy and Vissing-Jorgensen (2012) find that aggregate supply can explain a variety of yield spreads, and Greenwood and Vayanos (2014) analyze how the maturity structure of supply explains expected returns across the yield curve. Our contribution is to show similar effects in the equities market. Our focus on demonstrating broad, factor-level demand pressure is the key difference between this paper and the index-inclusion studies pioneered by Shleifer (1986). Admittedly, those studies have better identification, yet to achieve identification, they necessarily focus on rare events that impact individual stocks. Our study sheds light on demand pressures with broad-market impact.

Our paper also contributes to the literature that explains movements of stock style factors. Fama and French (1995) find evidence of factor structure in earnings but fail to find a strong link between the earnings factors and return factors. Barberis and Shleifer (2003) propose the “style investing” hypothesis: style-level comovement reflects investors’ tendency to think about stocks in categories, and positive-feedback trading lead to style-level reversals. Teo and Woo (2004) and Wahal and Yavuz (2013) document evidence consistent with this hypothesis. Our work differs from these papers because we document reversals that are generated by slow-moving liquidity provision, rather than mean-reverting demand pressures emphasized in those papers.

Our paper also adds to the literature linking fund flows and stock price movements. Frazzini and Lamont (2008) show that retail investors reallocate money to funds that subsequently underperform. Coval and Stafford (2007) show that fire sales induced by extreme mutual fund flows lead to large price impact and reversals, and Lou (2012) use fund flows to explain a number of short-run anomalies, notably momentum. We contribute to this literature by showing that fund flow induced price pressures are not only important at the individual stock level but also at the aggregate factor level. Finally, we also contribute to the burgeoning literature on the effect of slow-moving capital. Mitchell, Pedersen, and Pulvino (2007) document evidence of slow arbitrageur response to price dislocations in convertible bonds and merger arbitrage. Greenwood, Hanson, and Liao (2018) show the effect of slow-moving capital across markets, and Duflot (2010) reviews a number of other studies that can be interpreted as evidence for slow-moving capital.

The remainder of the paper is structured in three main sections. Section 2 documents the factor structure in fund flows, and shows that factor-level fund flows are associated with contemporaneous factor returns and subsequent price reversals. Under the guidance of a simple model, section 3 shows suggestive evidence for slow-moving liquidity provision. Section 4 uses a structural model to decompose the flow-driven component in returns and compare structurally estimated parameters to the existing literature. Section 5 summarizes our findings.
2 Price pressures of factor-level mutual fund flows

Most equity mutual funds in the U.S. are benchmarked to capitalization or valuation based indices, making them convenient vehicles for retail investors to reallocate investment across styles. This leads us to hypothesize that retail investors may use mutual funds to allocate investments across size and valuation-based styles, causing large style-level price impacts.

2.1 Data

We obtain mutual fund return and flow data from the CRSP survivorship-bias-free mutual fund database. Following Lou (2012), we restrict our attention to domestic equity funds (CRSP objective code starting with “ED”) that on average invest 75% to 125% of their portfolio in common stocks. We only use funds with more than $1 million assets under management (TNA).

There are two data choices worth explaining. First, we use annual observations to get the longest possible sample. We choose this because we need statistical power to detect time-series return predictability in factors. Annual observations start in 1962, while quarterly and monthly data start in 1970 and 1991, respectively. Our subsequent work shows that not using higher frequency data is not a major impediment because style-level flows are slow-moving, and the return reversal patterns span multiple years. Second, the CRSP data is at the share class level. Unlike some existing papers (e.g. Wermers (2000)), we do not merge data at the fund level because existing mappings do not have sufficient history. This, again, does not affect our analyses because our exercises are conducted at the factor-level, not at the fund-level. For simplicity, we refer to share classes as “funds” in subsequent writing.

While mutual fund data exists as far back as 1962, we use data from 1965 on because we need three years of data to estimate fund style loadings. We end up with 14,054 unique funds (share classes) and 456,261 fund-year observations. Comparison to the Federal Reserve flow of funds data shows that our sample covers over half of the U.S. equity mutual fund industry. Table I presents summary statistics.

2.2 Measuring fund styles

We measure fund styles using time-series regressions on Fama-French factors. For each fund $i$ at each December month $t$, we regress its trailing 36 monthly returns on the Fama-French three factors, obtained from Ken French’s website:

$$\text{Ret}_{i,t} = \alpha_{i,t} + \beta_{i,t}^{\text{MKT}} \cdot \text{MKT}_t + \beta_{i,t}^{\text{SMB}} \cdot \text{SMB}_t + \beta_{i,t}^{\text{HML}} \cdot \text{HML}_t + \epsilon_{i,t},$$

$3$Morningstar, the largest mutual fund analysis company, categorize funds into a three by three style box using size and value characteristics.

<table>
<thead>
<tr>
<th>Period</th>
<th>Fund-years</th>
<th>Funds</th>
<th>TNA (million $)</th>
<th>TNA Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965-1970</td>
<td>687</td>
<td>136</td>
<td>27,885</td>
<td>244</td>
</tr>
<tr>
<td>1971-1975</td>
<td>987</td>
<td>247</td>
<td>36,513</td>
<td>185</td>
</tr>
<tr>
<td>1976-1980</td>
<td>1,173</td>
<td>247</td>
<td>31,956</td>
<td>136</td>
</tr>
<tr>
<td>1981-1985</td>
<td>1,274</td>
<td>287</td>
<td>54,623</td>
<td>214</td>
</tr>
<tr>
<td>1986-1990</td>
<td>1,969</td>
<td>513</td>
<td>143,427</td>
<td>364</td>
</tr>
<tr>
<td>1991-1995</td>
<td>3,466</td>
<td>880</td>
<td>382,692</td>
<td>552</td>
</tr>
<tr>
<td>1996-2000</td>
<td>9,002</td>
<td>2,935</td>
<td>1,701,773</td>
<td>945</td>
</tr>
<tr>
<td>2001-2005</td>
<td>23,196</td>
<td>6,615</td>
<td>2,643,091</td>
<td>570</td>
</tr>
<tr>
<td>2006-2010</td>
<td>31,505</td>
<td>8,457</td>
<td>3,502,951</td>
<td>556</td>
</tr>
<tr>
<td>2011-2015</td>
<td>38,728</td>
<td>9,719</td>
<td>5,161,854</td>
<td>666</td>
</tr>
</tbody>
</table>

Table 1. Summary statistics. This table summarizes the domestic equity mutual funds obtained from the CRSP survivorship-bias-free mutual fund database. We restrict the sample to domestic equity mutual funds with an average of 75% to 125% of portfolio holdings devoted to equities over the sample period. A “fund” refers to a share class in this paper.

where \( \tau = t - 35, ..., t - 1, t \), and \( \text{MKT}_t, \text{SMB}_t, \text{HML}_t \) are the Fama-French factors. We require all past 36 monthly returns to exist. We then use the estimated size and value loadings (\( \hat{\beta}_{SMB}^{i,t} \) and \( \hat{\beta}_{HML}^{i,t} \)) to proxy for fund styles.

To ensure that factor loadings do capture fund styles, we compare them to Lipper fund objectives which is available in CRSP from 1998. Lipper classifies 32% of funds as having capitalization-based objectives (small cap, large cap, etc.) and 55% as having valuation-based objectives. Figure 2 plots the range of SMB loadings for cap-based funds and HML loadings for valuation-based funds. Reassuringly, funds that self-report to be more aligned with small-cap styles are associated with higher SMB loadings. There is also a clear correlation between HML loadings and valuation-based styles (growth versus income), although the relationship is slightly weaker, likely because the industry definition of growth and value differs from the Fama-French definition (Morningstar 2016).

An alternative way to infer fund styles is to estimate factor loadings of stocks, and then aggregate up the stock loadings over the fund portfolio holdings. This may be less noisy (Daniel and Titman 1997), but it relies on 13F holdings data which does not exist before 1980, while our sample starts in 1965. In unreported robustness tests, we find that inferring fund style using holdings after 1980 yield very similar results to our current method.
Figure 2. Distribution of factor loadings for funds with different style objectives. The left figure plots distribution of SMB (size factor) loadings for funds with capitalization-based Lipper fund objectives, and the right figure plots distribution of HML (value factor) loadings for those with valuation-based objectives. The top and bottom edge of the boxes are 25% and 75% percentiles and the whiskers are 5% and 95% percentiles. The sample starts from 1998 when Lipper objectives become available in the CRSP data set.

2.3 Size and value factors in mutual fund flows

We follow existing literature (e.g. Coval and Stafford (2007)) to measure fund flows by computing TNA (total net asset) changes that are not explained by fund returns. At each December month \( t \) and for each fund \( i \), we compute the annual dollar net fund flows using

\[
\text{DollarFlow}_{i,t-11 \rightarrow t} = TNA_{i,t} - TNA_{i,t-12} \cdot \prod_{s=t-11}^{t} (1 + \text{Ret}_{i,s})
\]  

where \( TNA_{i,t} \) is the total TNA at the end of month \( t \) and \( \text{Ret}_{i,t} \) is the post-fee fund return during month \( t \).

To detect flows at the style level, for each December month \( t \), we double sort funds using lagged factor loadings \( \hat{\beta}^{\text{SMB}}_{i,t-12}, \hat{\beta}^{\text{HML}}_{i,t-12} \) into five by five portfolios. This is the same procedure used by Fama and French (1996) who double sort stocks by market-capitalization and book-to-market characteristics. While we sort to have equal number of funds across portfolios in each year, Figure 11 in the appendix shows that the distribution of TNA is fairly even across the 25 portfolios, with large cap (low SMB loading) portfolios occupying slightly more weight.

For each portfolio \( p \) and each December month \( t \), we compute the portfolio-level flow after
normalizing by lagged total TNA:

\[
    \text{Flow}_{p,t-11\rightarrow t} = \frac{\sum_{\text{fund } i \in \text{portfolio } p} \text{DollarFlow}_{i,t-11\rightarrow t}}{\sum_{\text{fund } i \in \text{portfolio } p} \text{TNA}_{i,t-12}}.
\] (3)

We then use principal component analysis (PCA) to detect factor structures. Because we are interested in the flow into styles, we first subtract the annual average flow from each portfolio before applying PCA. This is akin to removing a “market factor” in flows.

If flows are idiosyncratic, then each principal component (PC) should explain \(1/24 \approx 4.2\%\) of variance, but we find a clear factor structure. The first two PCs explain 26\% and 18\% of variance, respectively. More interestingly, they appear to be linear combinations of size and value factors. As shown in Figure 3, the first PC portfolio is long small-value stocks and short large-growth stocks, while the second PC portfolio is long small-growth and short large-value.

Figure 3. Loadings of the first two principal components in the net flows to the 25 size-value mutual fund portfolios. In each year, we double sort funds into five by five portfolios using lagged SMB and HML factor loadings, and then compute the annual net inflow into each portfolio as a fraction of the total lagged TNA (equation (3)). We then conduct a principal component analysis on the flows of the 25 portfolios.

This finding is statistically robust. First, we make sure that the result cannot happen by chance due to a small sample. We bootstrap under the null hypothesis that portfolio flows retain their respective marginal distributions but are independent from each other. Under the null, the chance that the first two PCs explain so much variance is effectively zero (Figure 4). Second, in unreported robustness checks, we find qualitatively similar results when using quarterly data after 1970, or when using alternative ways to sort the size-value portfolios.

What explains this factor structure in flows? One natural hypothesis is the return-chasing channel. Fund flows chase returns (Chevalier and Ellison (1997)), and because returns have a factor structure (Fama and French (1996)), flows simply inherit that factor structure. Appendix B.2 shows that this is not the case. Specifically, we use panel regressions to control for fund flows that can be

---

Because we subtract the average annual flow, one degree of freedom is lost.
Figure 4. Fraction of flow variance explained by each principal component. The actual variance explained is marked using blue triangles. The black lines are the result when bootstrapping under the null hypothesis that flows in each of the 25 portfolios are independent. The black solid and dashed lines are 1%, 50%, and 99% percentiles from the bootstrapped results.

explained by past eight lags of quarterly returns, and then redo the PCA exercise using the flow residuals. The size and value factor structures remain, albeit slightly weaker: the combined variance explained by the first two PCs drop from 48.7% to 41.8 - 42.9%. [7]

Because return-chasing can only explain a small portion of the flow factor structure, this suggests that the factor-level flows arise from changes in investor preferences or shifts in beliefs about investment prospects of stock styles. Section 2.4 shows anecdotal evidence consistent with this. For instance, after the dot-com episode, we see very large reallocation of money from growth funds into value funds, consistent with investors being alarmed by the losses of internet growth stocks. [8]

2.4 Factor-level flows are large and positively correlated with factor returns

The fact that flows have factor structures immediately suggests that factor-level flows are large. Correlated flows do not cancel out when aggregating.

7 The raw flow-based results here are slightly different because we require the existence of lagged quarterly fund returns.

8 To what extent is the flow factors explained by return-chasing is not important for the subsequent analysis. We admit that it is always possible that our return-chasing model is misspecified, so we can end up underestimating the importance of return-chasing. The exact way in which flows chase past returns is subject to debate. For instance, [Chevalier and Ellison (1997)] suggest that the flow-performance relationship is convex while [Spiegel and Zhang (2013)] show it is linear. There is also evidence that flows do not chase raw returns but rather returns in excess of benchmark returns [Christoffersen and Simulin (2017)]. See [Christoffersen, Musto, and Wermers (2014)] for a review of this literature.
To examine flows at the factor level, we follow the Fama-French procedure to construct small-minus-big (SMB) and high-minus-low (HML) fund flow factors. In each year, we double sort funds into 2 by 3 bins using size and value loadings from the previous year, and define flow factors as:

\[
\text{Flow}^{\text{SMB}} = \frac{\text{Flow}_{(\text{small}, \text{high})} + \text{Flow}_{(\text{small}, \text{medium})} + \text{Flow}_{(\text{small}, \text{low})}}{3} - \frac{\text{Flow}_{(\text{big}, \text{high})} + \text{Flow}_{(\text{big}, \text{medium})} + \text{Flow}_{(\text{big}, \text{low})}}{3},
\]

\[
\text{Flow}^{\text{HML}} = \frac{\text{Flow}_{(\text{small}, \text{high})} + \text{Flow}_{(\text{big}, \text{high})}}{2} - \frac{\text{Flow}_{(\text{small}, \text{low})} + \text{Flow}_{(\text{big}, \text{low})}}{2},
\]

where the (small, high) refers to the portfolio of funds with small cap and high book-to-market characteristics, for instance.

Figure 5 plots the time series of flow factors against the corresponding return factors. The flow factors are very large, making it plausible for them to generate large price impacts. For example, investors sold out of the size factor over 1997 to 1999, and the overall reallocation totals $332 billion dollars, equaling 2.4% of the entire U.S. equity market capitalization at that time. After the dot-com episode in 2000, from 2001 to 2005, investors took money out of growth funds and moved it into value funds, with a total reallocation of $640 billion which is 4.1% of the entire market.

Consistent with large flows generating price impact, size and value factor flows are positively correlated with corresponding factor returns with 42% and 21% correlations, respectively. Figure 6 shows that the HML correlation is lower because of two outlier years - 1992 and 2000. In both cases, the return of value stock rebounded from a sudden crash, while the fund flows, being slow-moving in nature, continued to shift from growth to value funds. If those two years are excluded, the correlation for HML equals 44%, similar to that of SMB. Linear regressions of factor returns on factor flows also yield similar coefficients: 0.64 for SMB and 0.45 for HML, and the difference is not statistically significant with a t-statistic of 0.80.

### 2.5 Factor flows predict factor price reversals

If the factor-level fund flows create factor-level price pressures, we will expect price reversals. There is strong evidence of price reversal. Figure 1 in section 4 shows this graphically. We also run forecasting regressions for each factor Factor ∈ {SMB, HML}:

\[
\frac{\text{Ret}^{\text{Factor}}_{t+1-t+60}}{5} = a^{\text{Factor}} + b^{\text{Factor}} \cdot \text{Flow}^{\text{Factor}}_{t-11-t} + e^{\text{Factor}}_{t+1-t+60}
\]

where Ret^{\text{Factor}} is the log factor return, so the left hand side variables are annualized factor returns for subsequent five years. To adjust for overlapping time periods, we estimate standard errors following Hodrick (1992).

The results are reported in Table 2. The regression coefficients are statistically significant at 1% level for both style factors, and they are not statistically different from each other, with the t-stat of difference equaling 1.09. Moreover, the relationship is economically large: one standard deviation
higher flow (9.0% for SMB and 14.5% for HML) leads to lower annual factor return of 3.0% for SMB and 2.7% for HML in the subsequent years. The $R^2$ of 17.4% for SMB and 26.2% for HML are also large.

3 Slow-moving liquidity provision

Collectively, the empirical findings in section 2 suggest that factor-level mutual fund flows create price pressures that revert in subsequent years. We interpret the data through an illustrative model in which exogenous mutual fund flows are accommodated by slow-moving liquidity providers. The key feature in the model is that some liquidity providers, such as pension funds and hedge funds,
Figure 6. Scatter plot of contemporaneous factor returns and flows. Some years with extreme observations are marked.

<table>
<thead>
<tr>
<th></th>
<th>SMB</th>
<th>HML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.034*</td>
<td>0.044**</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Factor Flow</td>
<td>-0.337***</td>
<td>-0.189***</td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
<td>(0.069)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.174</td>
<td>0.262</td>
</tr>
<tr>
<td>N</td>
<td>46</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 2. Using factor flows to forecast subsequent factor returns. Standard error calculation follows Hodrick (1992).

are slow to respond to the flow shocks.

Because the flow and return reversal patterns last years, it is also natural to hypothesize alternative explanations with frictionless liquidity provision, the standard assumption in existing models (e.g. Barberis and Shleifer (2003)). We show that the key distinguishing feature of slow-moving liquidity is that price reversals are not accompanied flow reversals. In contrast, in models without slow-moving liquidity, flow reversal is necessary. Empirically, the data supports the slow-moving liquidity interpretation.

3.1 A model with slow-moving liquidity provision

This is a simplified adaptation of the model in Duffie (2010). There is a single risky asset being traded at discrete times $t = \ldots, -2, -1, 0, 1, 2, \ldots$. In our setting, we think of the asset being a style factor.
portfolio such as the Fama-French size portfolio. In each period, the asset pays independent and identically distributed dividend $D_t$ with distribution $\mathcal{N}(0, \sigma_D^2)$. Borrowing and lending is available at the exogenously specified gross risk-free rate $R_f = 1 + r_f > 1$.

There are two types of agents. The first type is noise traders with an aggregate exogenous asset demand process with level $Z_t$ in period $t$. We think of them as representing the mutual fund investors in our data, creating the factor-level flows, defined as changes in the exogenous demand levels: $f_t = Z_t - Z_{t-1}$. To ensure model stationarity, we assume flows follows a mean-reverting process $f_{t+1} = -\delta \cdot Z_t + \epsilon_{t+1}^f$ with mean-reversion parameter $\delta \in (0, 1)$ and noise $\epsilon_{t}^f$ independent and identically distributed with distribution $\mathcal{N}(0, \sigma_f^2)$.

The other type of agents is a unit mass of competitive liquidity providers with infinite-horizon time-separable constant absolute risk aversion utility. We think of these as representing all market participants other than mutual fund investors. To capture slow-moving liquidity provision, we assume that $q < 1$ fraction of them are “infrequent agents” who only trade their portfolios every $k \geq 2$ periods, with $1/k$ of them (thus a mass of $q/k$) adjusting in each period. When infrequent agents are not trading, their dividend income is reinvested at risk-free rate until their next investment decision. We think of the infrequent agents as slow-moving market participants such as pension funds. The speed at which they trade against flow-induced price movements does not reflect the frequency of portfolio rebalancing, but rather the speed at which they become aware of the price dislocations.

The remaining $1 - q$ mass of agents is made up of “frequent agents” who re-optimize their portfolios every period. We think of these as investors that actively track mutual fund flows or order flows, such as professional market makers or hedge funds. They are quick to learn about and profit from price dislocations. Note that this model reduces to the standard models with frictionless liquidity provision when $q = 0$.

Equilibrium is characterized by agent portfolio optimization and market clearing. Duffie (2010) shows that, under certain parameter restrictions, the model has a solution where asset price and agent demands are all linear in the state vector $Y_t \equiv (D_t, Z_t, x_{t-1}, x_{t-2}, ..., x_{t-k+1})$ where $x_t$ is the number of shares held by infrequent agents in period $t$. The positions held off-market show up because they affect the current aggregate security supply. The solution is obtained numerically.

**Discussion: the slow-moving liquidity assumption**

We emphasize that the slow response of infrequent traders in our model does not reflect the frequency of portfolio adjustment, but the speed at which investors become aware of the factor-level price dislocations. Market participants cannot trade against the flow-induced dislocations until they become aware of the style-level mutual fund flows which are not directly observable. While investors can find a lot of information about mutual funds online – past performance, investment style, manager profile, etc. – fund flows simply aren’t among them. To learn about style-level fund flows, one needs to subscribe to a database such as CRSP, compute fund-level flows, and then aggregate them at the factor level.
In our private conversations with ten fund managers and investment professionals, only one track fund flows, and none are aware of the large factor-level reallocations we document. They are all “infrequent agents” in our model. Those who do not track fund flows will eventually be driven to trade when they realize that price levels seem inconsistent with their estimate of fundamentals, but that action will always be lagged as the signal to noise ratio is low.

The likely candidates for “frequent agents” are broker-dealers who observe order flows as part of their business as well as hedge funds. These players are more likely to be quick to respond to mutual fund flows, but because they only hold a small fraction of the market, their trades have limited impact on prices. According to Federal Reserve Flow of Funds, during our sample period, security dealers only held an average of 0.38% of the U.S. equity market. Hedge funds became active in 1990s.\textsuperscript{9} To size the influence of hedge funds, we use Thomson Reuters Lipper Hedge Fund database to estimate the asset-under-management of hedge funds\textsuperscript{10} and then multiply by the 1.6 times average leverage estimated in Ang, Gorovyy, and Van Inwegen (2011). After applying leverage, we estimate that hedge fund equity investment as a fraction of overall U.S. equity market is an average of 0.74% since 1990 and peaks at 1.83% in 2007, right before the financial crisis. Thus, the combined capital of security dealers and hedge funds is not always enough to absorb the large mutual fund flows we document in section 2.4.

3.2 Testing for slow-moving liquidity

Most existing models with heterogeneous agents, such as Barberis and Shleifer (2003), assume frictionless liquidity provision. Our model can also capture frictionless liquidity provision by setting the fraction of infrequent trader $q=0$, a case that we solve in appendix A.\textsuperscript{11}

To illustrate the key feature of slow-moving liquidity, Figure 7 plots price paths after a one-time, permanent supply shock. With frictionless liquidity as in existing models, asset price level immediately drops to a new equilibrium level with no reversion. In contrast, with slow-moving liquidity, because only a subset of liquidity providers are present to absorb the supply shock (mass of $(1 - q) + \frac{q}{k}$), asset price overshoots initially. As more infrequent agents arrive to share risk, the temporary price movements gradually dissipates and price level reverts towards the new equilibrium. While the standard model only generates permanent price impact, the slow-moving liquidity model features amplified volatility due to transitory movements.\textsuperscript{12} We formalize this difference in a testable prediction.

**Proposition 1.** *(Testing for slow-moving liquidity provision)*

*If there is no slow-moving liquidity provision ($q=0$), then after controlling for future flows, current*\textsuperscript{9}See article “Roaring ahead – Exchange-traded funds have overtaken hedge funds as an investment vehicle” by The Economist, accessible at https://www.economist.com/finance-and-economics/2015/08/01/roaring-ahead\textsuperscript{10}We use funds with primary investment objectives of “equity market neutral” and “long/short equity” to estimate investment in equity strategies.\textsuperscript{11}The other models differ in other aspects, especially in the mechanism behind generating the flow shocks.\textsuperscript{12}The long-run equilibrium price change is also larger in the slow-moving capital. This is also due to slow-moving capital which increases non-fundamental risk, so the quantity of price risk per share of asset is higher.
Figure 7. The price path following a large supply shock in different models. The frictionless liquidity model refers to the case where all liquidity providers respond immediately ($q = 0$). In the slow-moving liquidity model (solid line), fraction $q = 80\%$ of traders are “infrequent agents” that participate every $k = 8$ periods. Other parameters: $\delta = 0.05, \text{Var}(D_t) = \text{Var}(Z_t) = 0.1, \gamma$ (agent risk aversion) = 1., and $\tau_f = 5\%$.

Flows do not forecast future price movements. Concretely, for any horizon $h \geq 1$,

$$E(\underbrace{P_{t+h} - P_t}_{\text{price reversion}} | \underbrace{f_t, f_{t+1}, \ldots, f_{t+h}}_{\text{future flows}}) = E(\underbrace{P_{t+h} - P_t}_{\text{future flows}} | \underbrace{f_{t+1}, \ldots, f_{t+h}}_{\text{future flows}}).$$

(7)

In contrast, with slow-moving liquidity provision, flows predict future price reversals after controlling for future flows. The proof of Proposition 1 is immediate from the solution of the $q = 0$ model in appendix A.

Existing models with frictionless liquidity provision in the literature generate price reversions through reversion of demand shocks. For instance, in the model of Barberis and Shleifer (2003), the demand of “switchers” is mean-reverting. In contrast, the slow-moving liquidity model generates price reversions that are not generated by mean-reverting flows.

To implement the test of Proposition 1, we run the following return forecasting regression for each factor:

$$\frac{\text{Ret}^{\text{Factor}}_{t+1 \rightarrow t+60}}{5} = a + b \cdot \text{Flow}^{\text{Factor}}_{t-11 \rightarrow t} + c_1 \cdot \text{Flow}^{\text{Factor}}_{t+1 \rightarrow t+12} + \ldots + c_5 \cdot \text{Flow}^{\text{Factor}}_{t+49 \rightarrow t+60} + \text{Flow}^{\text{Factor}}_{t+1 \rightarrow t+60}.\text{flows in subsequent five years}$$

(8)

Specifically, switchers chase recent price changes with exponentially decaying weights. When good fundamental shocks hit, prices jump up, and switchers buy more of the asset. As the upward price movement becomes farther away in history, the optimal holding level of switchers reverts back and thus they sell. This gives rise to mean-reverting demand.
If liquidity provision is frictionless, coefficient $b$ should become zero after controlling for future flows. However, results in Table 3 show otherwise. The return forecasting power remains robust. Further, Figure 12 in the appendix shows that flows are positively autocorrelated, not mean-reverting.

<table>
<thead>
<tr>
<th>Subsequent five year factor return (annualized)</th>
<th>SMB</th>
<th>HML</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.007</td>
<td>0.034*</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Flow</td>
<td>-0.376***</td>
<td>-0.337***</td>
</tr>
<tr>
<td></td>
<td>(0.115)</td>
<td>(0.117)</td>
</tr>
<tr>
<td>Subsequent Flow, 1 year</td>
<td>0.125**</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>Subsequent Flow, 2 year</td>
<td>0.115</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.088)</td>
</tr>
<tr>
<td>Subsequent Flow, 3 year</td>
<td>0.099*</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>Subsequent Flow, 4 year</td>
<td>0.014</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>Subsequent Flow, 5 year</td>
<td>0.380***</td>
<td>0.131*</td>
</tr>
<tr>
<td></td>
<td>(0.133)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.481</td>
<td>0.174</td>
</tr>
<tr>
<td>N</td>
<td>46</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 3. Using factor flows to forecast subsequent factor returns. Specifications (1) and (3) control for flows in subsequent five years. (2) and (4) are the same results in Table 2, reported here for convenience. Factor-level flows are constructed using the procedure described in section 2.4. Standard errors calculation follows Hodrick (1992).
4 Decomposing flow-induced factor price variation

In this section, we use structural vector autoregressive models to capture the dynamics of factor flows and returns. This allows us to isolate the component of flow-induced factor returns. Also, we can estimate the price impact coefficients and compare them to those found in the existing literature.

We fit two separate models for each factor

\[
\begin{pmatrix} R_t \\ F_t \end{pmatrix} = \begin{pmatrix} \mu_R \\ \mu_F \end{pmatrix} + \begin{pmatrix} \lambda \cdot F_t \\ \beta \cdot R_t \end{pmatrix} + \begin{pmatrix} B_{1,RR} & B_{1,RF} \\ B_{1,FR} & B_{1,FF} \end{pmatrix} \begin{pmatrix} R_{t-12} \\ F_{t-12} \end{pmatrix} + \ldots + \begin{pmatrix} B_{5,RR} & B_{5,RF} \\ B_{5,FR} & B_{5,FF} \end{pmatrix} \begin{pmatrix} R_{t-60} \\ F_{t-60} \end{pmatrix} + \begin{pmatrix} u_{R,t} \\ u_{F,t} \end{pmatrix},
\]

where \( R_t \) and \( F_t \) denote the annual log return and flow in the year ending in December month \( t \).

In our specification, there are two effects contributing contemporaneous correlation between flows and returns. \( \lambda \) is a price impact coefficient and captures the impact of flows on returns. \( \beta \) captures flows chasing returns within the same year. Prior papers find that mutual fund flows chase returns using monthly or quarterly data (Chevalier and Ellison (1997), Coval and Stafford (2007)), which can show up as a contemporaneous relationship in our annual sample. To separate the impact of these two forces, we estimate \( \beta \) using monthly data so that the model is identified.

4.1 Estimate return-chasing parameter \( \beta \) using monthly data

We decompose annual log flows and log returns in 12 months:

\[
\begin{align*}
F_t &= \text{Flow}_t + \text{Flow}_{t-1} + \ldots + \text{Flow}_{t-11}, \\
R_t &= \text{Ret}_t + \text{Ret}_{t-1} + \ldots + \text{Ret}_{t-11}.
\end{align*}
\]

Suppose that, at the monthly frequency, flows chase past returns in a linear model:

\[
\text{Flow}_t = a + b_0 \cdot \text{Ret}_t + b_1 \cdot \text{Ret}_{t-1} + b_2 \cdot \text{Ret}_{t-2} + \ldots + b_H \cdot \text{Ret}_{t-H}
\]

flows chasing returns

\[
+ c_1 \cdot \text{Flow}_{t-1} + c_2 \cdot \text{Flow}_{t-2} + \ldots + c_H \cdot \text{Flow}_{t-H} + \epsilon_t.
\]

autocorrelation in flows

where \( H \) is a long enough horizon to capture all the flows. Then, the return-chasing coefficient of the annual variables can be expressed as a function of the monthly coefficients. Assuming that returns are not serially correlated and homoskedastic, we have:

\[
\beta = \frac{\text{Cov}(F_t, R_t)}{\text{Var}(R_t)} = \frac{\text{Cov} (\text{Flow}_t + \ldots + \text{Flow}_{t-11}, \text{Ret}_t + \ldots + \text{Ret}_{t-11})}{\text{Var} (\text{Ret}_t + \ldots + \text{Ret}_{t-11})}
\]

\[
= \frac{12 \cdot b_0 + 11 \cdot b_1 + \ldots + 2 \cdot b_{10} + b_{11}}{12}
\]

(13)
By attributing the contemporaneous relationship $b_0$ to return-chasing, we err on the side of over-estimating the effect of return chasing, and thus underestimate the price impact of flows. To get a statistically precise estimate of the coefficients in (12), we use monthly data of all mutual fund shares since 1990 in CRSP, with a total of 1,015,653 fund-month observations. We run the panel regression version of (12) and cluster standard errors by month and fund. We use $H = 24$ lags because Barber, Huang, and Odean (2016) find that 18 months is roughly sufficient to capture the flow response to past returns. We weight observations by fund TNA because larger funds are more important in our size and value flow factors. The regression coefficients plotted in Figure 8 confirm that 24 months is sufficient to capture the dynamics.

Figure 8. Estimated coefficients in return-chasing regression. For all fund (share class) $i$ in month $t$, we run TNA-weighted regression:

\[
\text{Flow}_{i,t} = a + b_0 \cdot \text{Ret}_{i,t} + b_1 \cdot \text{Ret}_{t-1} + b_2 \cdot \text{Ret}_{t-2} + \ldots + b_H \cdot \text{Ret}_{t-H} + \ldots + \cdot b_{24} \cdot \text{Ret}_{t-24} + \ldots + c_{24} \cdot \text{Flow}_{t-24}
\]

and cluster standard errors by month and fund. The left figure plots coefficients capturing the return-chasing behavior of flows ($b_0, b_1, \ldots, b_{24}$), and the right figure plots coefficients capturing the autocorrelation of flows ($c_1, c_2, \ldots, c_{24}$). The dashed lines are two standard error bands.

Using (13), we estimate $\beta$ to be 0.124. The standard error calculated using the Delta method is 0.015. The estimation is relatively precise, so we plug in the point estimate in subsequent analysis to identify the SVAR model.
4.2 Reversion speed of flow-induced price movements

Having obtained a value for $\beta$, we now estimate the SVAR models for each factor. To visualize the impact of flows, in Figure 9, we plot the cumulative responses to a one standard deviation shock to the structural flow residual ($u_{F,t}$ in equation (9)).

Flow shocks lead to contemporaneous factor price movement that reverses in the subsequent two to five years. The slow reversion is partially due to the autocorrelation of flows, as shown in the lower panels of Figure 9. For the size factor, flows continue in the same direction for three more years, after which the return reversion takes hold. The flows only continue for one more year for the value factor, and the return reversion also happens faster for the value factor. Thus, if flows are not autocorrelated, we expect the reversion to complete in roughly one to three years.

This slow price reversion is consistent with two prior papers that investigate the reversion from mutual fund flow-induced price pressures. Coval and Stafford (2007) study the impact of flow-induced fire sales/purchases of mutual funds and observe price reversion that last at least 18 months (Figure 2 in their paper). Similarly, Figure 2 in Edmans et al. (2012) shows that mutual fund flow-induced price pressures take two years to revert.
4.3 Price impact coefficients

The point estimates of the price impact coefficient \( \lambda \) is 0.281 for size and 0.370 for value, with p-values of 0.17 and 0.04. While there is quite a bit of estimation error, we can be reasonably certain that our finding positive price impact of flows is not due to statistical errors. To calculate p-values, we bootstrap the SVARs by resampling the residuals under the null of \( \lambda = 0 \), and then re-estimate \( \lambda \). Figure 10 shows the distribution of \( \lambda \) estimates under the null.

Are these price impact coefficients reasonable? We now convert them into demand elasticities and compare them to estimates from well-identified index exclusion studies. Note that the returns in the SVARs are calculated using stock capitalization as the denominator, while the flows are calculated using mutual fund TNAs as the denominator, so we need to adjust for the difference.
Figure 10. Distribution of price impact coefficient estimates ($\hat{\lambda}$) under the null of $\lambda = 0$. To get these distributions, we bootstrap the SVAR models for size (SMB) and value (HML) after setting $\lambda = 0$, and then re-estimate $\lambda$. The dashed lines are points estimates in the actual data.

Concretely:

\[ \text{Elasticity} = \frac{\text{Flow/Total Value}_{\text{stock}}}{\Delta \text{Price}_{\text{stock}}/\text{Price}_{\text{stock}}} \]  \hspace{1cm} (16)

\[ \lambda = \frac{\Delta \text{Price}_{\text{stock}}/\text{Price}_{\text{stock}}}{\text{Flow/Total value}_{\text{mutual funds}}} \]  \hspace{1cm} (17)

Therefore, we need to make an adjustment:

\[ \text{Elasticity} = \frac{1}{\lambda} \cdot \frac{\text{Total value}_{\text{mutual funds}}}{\text{Total value}_{\text{stock}}} \]  \hspace{1cm} (18)

Over our sample period, mutual funds stock holdings represented on average approximately $1/6$ of the U.S. stock market (Fed flow of funds). Adjusting for this gives points estimates of demand elasticities equal to 0.59 using the SMB estimate and 0.45 for the HML estimate. We view this as a lower bound because the overall style-level retail flows may be significantly larger. If the mutual fund flows we estimate reflect changes in investor preferences over styles, then it is plausible that retail investors may also conduct similar style reallocation in their direct style holdings. Over our sample period, Fed flow of funds data shows that only around $1/4$ of household equity holdings (excluding pension, life insurance, etc.) is through mutual funds. If household investors conduct the same degree of style reallocation in their non-mutual fund holdings\(^{14}\), then we get demand elasticities of 1.80 to 2.37. We consider this an upper bound.

\(^{14}\text{That is, suppose for each 1}\% \text{ of reallocation into small-cap mutual funds, the “household” sector in Flow of Funds also reallocate 1}\% \text{ into small cap stocks through their other holdings.}\)
Our estimate is in the same range with that found in index inclusion studies. Using a large sample of Russell 1000 and Russell 2000 index redefinitions, Chang, Hong, and Liskovich (????) find an elasticity of 0.39 - 1.46. Using a smaller set of S&P 500 index changes, Shleifer (1986) find an elasticity of around 1.

4.4 Quantifying flow-driven return variation

How much factor return variance is caused by flow-induced price pressures and reversions? To estimate the counterfactual return variance without flows, we simulate the SVARs after making all flows and the structural flow-residuals equal to zero. On average, without the impact of flows, SMB annual return volatility reduces from 11.94% to 10.09%, and HML annual return volatility reduces from 13.92% to 11.94%. The amount of return variance explained by flows is thus 28.5% for SMB and 26.4% for HML.

This high explanatory power should not come as a surprise, as using flows to predict subsequent return reversals already gives \( R^2 \) in the 17% - 27% range (section 2.5). Because flows also cause the initial price impact in addition to the reversion, the overall \( R^2 \) is larger.

5 Summary

We show that style-level mutual fund flows create price pressure and reversions, and this flow-induced mechanism can explain close to 30% of Fama-French size and value factor return variance. Using mutual fund flows to measure retail investor reallocation of capital across stock styles, we find a strong factor structure in their order flows along size and value directions over the period of 1965 to 2015. The factor-level flows are in the order of hundreds of billions of dollars. Consistent with the flows creating temporary price impact, factor-level flows are contemporaneously correlated with factor returns, and predict subsequent factor price reversals with \( R^2 \) in the 17 - 27% range.

We argue that retail fund flows can create such large price impact because other investors (e.g. pension funds) are slow to provide liquidity. Guided by a model, we find evidence of slow-moving liquidity provision: price reversals are not accompanied by reversal of flows. This cannot be explained by “rational” or “behavioral” models that assume frictionless liquidity provision.

To quantify the flow-induced price variation, we fit structural vector-autoregressive models to capture the joint dynamics of flows and returns. The estimated price impact coefficients are similar to those in well-identified index inclusion studies, and the price reversion speeds are similar to those in existing literature on mutual fund flow-induced fire sales. Using the structural model, we find that flows explain close to 30% of total size and value factor return variance.

Overall, this paper provides evidence that demand shocks can explain a large fraction of stock market movements, and also highlights the importance of slow-moving liquidity provision in amplifying the price impact of demand shocks.
References


APPENDIX

A  Solving the model with frictionless liquidity provision \((q = 0)\)

As is typical in such models, we conjecture that the price is linear in the state variable:

\[
P_t = \lambda Z_t \quad (19)
\]

So positive exogenous demand raises prices. Merely for the purpose of expositional simplicity we assume agents do myopic portfolio optimization in derivations below. This is without loss of generality as Appendix C of \cite{Kozak, Nagel, Santosh (2017)} shows that the same pricing function is obtained when agents do long-horizon optimization.\(^{15}\)

Under the pricing conjecture \((19)\), return from \(t\) to \(t+1\) is given by

\[
R_{t+1} = P_{t+1} + D_{t+1} - R_f P_t
= \lambda Z_{t+1} + D_{t+1} - R_f P_t \quad (20)
\]

The conditional expectation and variance of returns given \(P_t\) are given by

\[
E_t(R_{t+1}) = \lambda E_t(Z_{t+1}) - R_f P_t
= \lambda(1 - \delta)Z_t - R_f P_t \quad (22)
\]

\[
Var_t(R_{t+1}) = Var_t(P_{t+1} + D_{t+1})
= Var_t(\lambda Z_t) + Var(D_{t+1})
= \lambda^2 \sigma_f^2 + \sigma_D^2 \quad (23)
\]

It is well known in such settings that mean-variance optimization gives the optimal demand curve

\[
x^*_t(P_t) = \frac{E_t(R_{t+1})}{\gamma Var_t(R_{t+1})} = \frac{\lambda(1 - \delta)Z_t - R_f P_t}{\gamma(\lambda^2 \sigma_f^2 + \sigma_D^2)} \quad (27)
\]

\(^{15}\) In particular, if we specialize their result to a single asset model like in our case, then they show that security price will be affine in the exogenous shock variable, just like in our case with myopic agents. There is one slight difference in specification. In their model, the exogenous shock is specified as an AR(1) process on the beliefs of a subset of agents, while we are assuming an AR(1) process on the exogenous demand itself.
The pricing conjecture in (19) is verified through market clearing:

\[ x_t^*(P_t) + Z_t = 0 \] (28)

\[ \Rightarrow P_t = \left( \frac{\lambda(1 - \delta) + \gamma(\lambda^2 \sigma_f^2 + \sigma_D^2)}{R_f} \right)^{\text{need to } \lambda} Z_t \] (29)

Thus, \( \lambda \) is given by solving a quadratic equation which has real solutions if flow and dividend shocks are not too volatile, or if risk aversion is not too high\(^{16}\). When there are two positive roots, we pick the one implying lower price volatility following prior literature \( \text{Bogousslavsky (2016)} \).

\(^{16}\)Specifically, the condition for having real solutions is:

\[ \delta + r_f \geq 2\gamma \sigma_f \sigma_D \] (30)
B Additional empirical results

B.1 Data description

---

**Figure 11.** Fraction of TNA for the 25 size-value sorted PCA portfolios. Blue lines are 25%, 50%, and 75% percentiles of the fraction of TNA in each PCA portfolio, and the green dashed lines is the average (1/25 = 4%). For example, portfolio “s1h2” refers to the one with lowest SMB loading (so corresponding to large cap) and second lowest HML loading.

---

**Figure 12.** Autocorrelation of factor flows. The blue dashed lines are 95% confidence intervals.
B.2 Is the factor structure in flows due to factor structure in returns?

We decompose the part of fund flows attributable to chasing past fund returns using a panel regression. We define return of fund \( i \) in quarter \( q \) as

\[
Flow_{i,q} = \frac{TNA_{i,q} - TNA_{i,q-1} \cdot (1 + Ret_{i,q})}{TNA_{i,q-1}}.
\]

We then run a panel regression on past fund returns, as well as time and fund fixed effects:

\[
Flow_{i,q} = a + b_1 \cdot Ret_{i,q-1} + \ldots + b_8 \cdot Ret_{i,q-8} + \sum_q \gamma_{\text{quarter}}^q \cdot 1_{\text{quarter } q} + \sum_q \gamma_{\text{fund}}^q \cdot 1_{\text{fund } i} + \epsilon_{i,q}.
\]

We choose eight lags because prior work indicate this is approximately sufficient to capture return-chasing (e.g. Barber et al. (2016) use 18 months, or 6 quarters). We then take the flow residuals in this regression, aggregate to annual frequency, and redo the principal component analysis exercise. Because flows (as a fraction of TNA) return more strongly to returns in smaller funds, we also do another version where we run regression (32) separately for funds in different size quintiles. Because we require having quarterly fund flow data for this exercise, the sample starts from 1973 and is smaller than that in the main paper which starts from 1965.

Table 4 shows the amount of flow variance explained by the first three PCs. There is still clearly a factor structure. After purging the impact of return chasing, the importance of the first PC declines slightly, but not much, from explaining 29.26% variance to 22.92%. The importance of the second PC does not change.

<table>
<thead>
<tr>
<th>Fraction of Variance explained by PC</th>
<th>Fund flows</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC</td>
<td>Original using returns</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>1</td>
<td>29.26%</td>
</tr>
<tr>
<td>2</td>
<td>19.45%</td>
</tr>
<tr>
<td>3</td>
<td>6.81%</td>
</tr>
<tr>
<td>4</td>
<td>6.24%</td>
</tr>
<tr>
<td>5</td>
<td>5.45%</td>
</tr>
</tbody>
</table>

Table 4. Marginal $R^2$ of PCs for fund flows not explained by return-chasing. Column 1 uses the original fund flows, and column 2 uses the fund flow residuals after a panel regression on past 8 lags of quarterly fund returns (equation (32)). Column 3 is like column 2 except that the regression on past fund returns is conducted separately by fund size quintiles.

Table 5 shows the loadings of the first two principal components. We use conditional coloring to make the pattern easier to see. The first two PCs are still linear combinations of size and value, although the first and second PC swapped. After taking out the impact of return chasing, the first
PC now becomes the “long small-growth, short big-value” portfolio, which was the second PC in the raw flows. Taken together, these results indicate that the factor structure in fund flows is unlikely to be explained away mechanically by return-chasing.

Table 5. Loadings of PCs for fund flows not explained by return-chasing. We apply the same principal component analysis procedure as in section 2.3, but in Panel B, we use flow residuals after regressing out the impact of past eight quarterly fund returns. In Panel C, we regress out the impact of past fund returns separately for funds in different TNA quintiles. Due to requiring quarterly flows data, this sample starts in 1973 and is slightly different from the annual sample starting in 1965 in the main paper.