Fund Flows, Slow-Moving Liquidity Provision, and Common Factors in Stock Returns

Jiacui Li*
Graduate School of Business
Stanford University

August 3, 2017

*I am grateful for comments from Nick Barberis, Simcha Barkai, Jonathan Berk, Svetlana Bryzgalova, John Cochrane, Darrell Duffie, David Hirshleifer, Lunyang Huang, Arvind Krishnamurthy, Hanno Lustig, Tim McQuade, Markus Pelger, Cameron Peng, Ken Singleton, Yang Song, and Eric Zitzewitz. All errors are mine. Please address correspondence to: Jiacui Li, Stanford GSB, 655 Knight Way, Stanford, CA 94305. E-mail: jiacui@stanford.edu.

Fund Flows, Slow-Moving Liquidity Provision, and Common Factors in Stock Returns

Abstract

Over the period of 1965-2015, retail investors frequently made large and uninformed capital reallocations at size and value style levels by trading equity mutual fund shares. Consistent with other market participants being slow to provide liquidity, fund flows are associated with large contemporaneous price impact which reverses in the subsequent years, explaining approximately 30% of SMB and HML factor return variance. Because price reversions are not accompanied by flow reversions, the evidence is inconsistent with standard "rational" or "behavioral" models with heterogeneous agents. Without slow-moving liquidity provision, flows would not be able to explain such a high fraction of broad-market price movements over long periods of time.

1 Introduction

A central question in financial economics is whether demand shocks that are unrelated to cash flows can create large stock price movements. Since the seminal work of Shleifer (1986), papers examining the effect of stock index redefinitions or the effect of uninformed mutual fund flows (e.g. Lou (2012)) have fairly convincingly demonstrated the existence of an effect. However, to achieve clean identification, these papers zoomed in on relatively narrow settings, which may leave readers wondering how large a fraction of price movements can be explained by demand shocks. Metaphorically, prior papers show that the price impact of demand shocks have "high t-statistic", but not necessarily "high R-squared". There is also a theoretical issue with the mechanism: when applying usual assumptions about preferences of investors who provide liquidity to demand shocks, empirically observed flows cannot account for a high fraction of price movements, unless we assume risk aversion parameters much higher than that estimated from microeconomic or experimental studies.[1]

This paper fills these two gaps. First, we empirically show that around 30% of the Fama-French SMB and HML factor return variance over the last 50 years arises from uninformed, large retail investor reallocations at the style level. Secondly, we hypothesize and show evidence for slow-moving liquidity provision amplifying and prolonging flow-induced price dislocations. The idea is, when mutual fund flows hit the market, only a subset of the most vigilant market participants are present to absorb them, so prices move excessively at first, and only gradually revert towards the new long-term equilibrium level as other investors also rtade against the price dislocation. So in addition to creating a permanent change in the equilibrium price level due to liquidity provider risk aversion, flow shocks also create transitory, mean-reverting price movements where reversion speed is governed by how quickly liquidity providers respond.

We start by documenting a strong factor structure in mutual fund flows along size and value styles. Just like how Fama and French (1996) sort stocks, we double sort funds by their SMB and HML loadings into 5 by 5 portfolios, and examine the correlation structure of fund flows in the

[1] For instance, Greenwood and Vayanos (2014) found that, if one assumes that mutual funds, pension funds, and insurance companies all participate in absorbing government bond supply, then they need a relative risk aversion coefficient of 91.2 to account for empirically observed price impact caused by government bond supply shocks. This is in contrast to the estimates from micro data which frequently gives values around 2, such as in Chetty (2006).

25 portfolios. Principal component analysis reveals a strong factor structure. The first principal component is akin to a "market" factor, and the next two components, explaining 16% and 11% variance, respectively, appear to be linear combinations of SMB and HML factors. One may doubt that the factor structure in flows may inherits the factor structure in returns due to the fact that fund flows chase returns, but we show that the flow factor structure persists after removing the part of return-chasing fund flows from data.

The style-level fund flows are large enough to generate meaningful price movements even in standard neoclassical models. For example, from 1997 to 1999, retail investors reallocated $332 billion from small cap funds to large cap funds, which represented 2.4% of total equity market value; during 2001 to 2005, they rotated $640 billion from growth funds to value funds, which was 4.1% of total market value. As documented in Lou (2012), fund managers respond to these fund flows by expanding or reducing holdings in their designated stock styles, so the large retail fund flows translate to large buying and selling pressures on the underlying stocks.

The style level fund flows are not informative about future cash flows, but are associated with large contemporaneous factor price movements and strong reversions in subsequent years. For both SMB and HML, a one standard deviation higher (or lower) fund flow is associated with 2-3% lower (or higher) annual factor returns in the 5 years to come. Using standard long-horizon return predictability regressions, we find the reversions are statistically significant for both factors with R^2 around 20%. We argue the price reversion is most consistent with slow-moving liquidity provision as in Duffie (2010)'s slow-moving capital model in which liquidity providers react to demand shocks gradually over time. This is plausible because most market participants, even though they may be willing to trade against such flows, do not monitor fund flows on a regular basis.[2]

There are two natural alternative hypotheses for the patterns we document. One explanation is that our findings arise mechanically from exogenous mean reversion in factor prices and return-chasing fund flows. Even though flows likely chase returns with a lag, when measured at a less granular frequency, this can manifest as "contemporaneous price impact". And because factor prices mean revert, due merely to their comovement with factor returns, flows will also appear to predict price reversion. However, under this story, reversion predictability should disappear after first

[2]In our private conversations with asset managers regarding our paper, we found that they do not keep track of mutual fund flows on a regular basis, and all except one are unaware of the existence of the large factor-level flows we document.

residualizing flows against current and past factor returns, but that is not supported by data.

The more challenging question is whether our evidence can be reconciled with a neoclassical model where investors have heterogeneous and mean-reverting preferences or beliefs. Under this story, when a preference or belief shock hits and makes retail investors more risk averse about certain stocks, they will sell to other investors, and this will generating flows in equilibrium. Because the overall demand for these stocks in the economy decreases, prices go down contemporaneously. If preferences of retail investors are mean-reverting, then on average they will buy some shares back later, and prices will revert back when they do so. This is the central mechanism behind many existing "rational" models where flows are generated by exogenous preference or endowment shocks, or "behavioral" models where flows come from distorted beliefs or some mechanical portfolio behavior.

We show that the evidence is inconsistent with this neoclassical view. The key difference between these neoclassical models and the slow-moving capital model lies in the joint dynamics of flows and price reversion. In the neoclassical framework, price reversion is driven purely by mean-reverting flows between the retail and non-retail sectors, but that is not true in the slow-moving capital model. If the neoclassical framework is right, then after controlling for future flows, we should cease to find current flows predicting future price reversals. That is unsupported in the data so we take this as a strong case against the neoclassical interpretation. Of course, we only reject this one class of neoclassical models – which we argue is representative of effectively all models in the prior literature – but it is possible that our evidence can be accommodated by models with more flexible cash flow processes and preferences.

How much factor return variation can be accounted by flows and slow-moving liquidity provision? We fit a structural vector autoregressive model to the joint dynamics of factor flows and factor returns over 1965 to 2015, and find that a high fraction – around 30% – of annual SMB and HML factor return variation is caused by flows. This high explanatory power should not come as a surprise as the reversion forecasting regressions already generate R^2 around 20%; the total impact on returns also includes the initial price impact. Reassuringly, the estimated price impact coefficient is similar to that found in the index inclusion literature, and the speed of price reversion is also similar to that found by other papers examining reversions after fund flow-induced price dislocations.

This paper is related to three literatures. First, in terms of demonstrating the demand impact on stock prices, there is the index inclusion literature which shows stock index composition changes are associated with significant price movements[3]. There are also a number of studies that also examine fund flow-induced price effects at individual stock levels (Lou (2012), Frazzini and Lamont (2008)), in fire-sale cases (Coval and Stafford (2007)), or during financial crises (Cella, Ellul, and Giannetti (2013)). As mentioned earlier, the key difference of this paper is in demonstrating that demand effects can explain a high fraction of market-wide movements over long periods of time. Secondly, this paper is related to a large number of papers trying to address the behavior of Fama-French factors . The relationship is loose, as existing papers almost always take the variation of size and value factors as given and try to address why SMB and HML have positive expected returns, while we explain why the factors are volatile to begin with.[4] Thirdly, following Duffie (2010), there is burgeoning recent literature empirically looking at the impact of slow-moving capital (e.g. Greenwood, Hanson, and Liao (2016)), to which our paper contributes by looking at slow-moving capital in accommodating factor-level flows.

The rest of the paper is as follows. Section 2 presents the slow-moving capital model and highlights how it generates price reversions through a mechanism different from standard neoclassical models. Section 3 describes the data, demonstrates the factor structure in flows, argues that flows are not informative about cash flows, and show price impact and price reversion patterns consistent with the slow-moving capital model. Section 4 shows that data is inconsistent with neoclassical and some other alternative hypotheses. Section 5 fits a structural model to separate the component of flow-driven return variation, estimates price impact coefficients and reversal patterns, and compare them against estimates from the existing literature. Section 6 concludes.

[3]This literature started from Shleifer (1986); see Chang, Hong, and Liskovich (2014) for a relatively recent contribution.

[4]Fama and French (1995) finds that the factor returns are not tightly linked to factor-level cash flow innovations, so there is an "excess volatility puzzle" in size and value factors, which we address.

2 Theoretical Framework

We interpret the data through a simplified version of the slow-moving capital model in Duffie (2010). The key departure of this model from the neoclassical framework is that some traders do not immediately respond to investment opportunities. In our application, the investment opportunities are the potential profits from trading against style-level price dislocations created by uninformed fund flows. In general, capital can be less than perfectly mobile due to a number of frictions, but we hypothesize that the slow response in our setting is, at least in a large part, coming from the fact that style-level flows are hard to observe. While we can find lots of information about mutual funds online – past performance, investment style, manager profile, etc. – fund flows simply are't among them, and one needs to subscribe to a database like CRSP and back out flows by computing AUM changes that are not explained by fund returns. One also needs to aggregate the individual fund flows up to style levels to observe the factor-level flows we use in this paper. In our conversation with over ten fund managers and investment professionals, all but one do not track fund flows at all, and none are aware of the large factor-level reallocations we document. 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 will almost surely come with a significant lag as the signal to noise ratio is low.

2.1 A model with slow-moving liquidity provision

At discrete time periods $t = ..., -2, -1, 0, 1, 2, ...$, agents submit demand curves to trade a single asset which can represent, for instance, the SMB portfolio. In each period the asset pays independent and identically distributed dividends distributed as $D_t \sim \mathcal{N}(0, \sigma_D^2)$. Borrowing and lending is infinitely available at exogenous gross risk-free rate $R_f = 1 + r_f > 1$.

There is a mass 1 of agents with infinite-horizon time-separable constant absolute risk aversion utility. Let $\gamma > 0$ denote their risk aversion parameter. Of these agents, $q < 1$ fraction are "infrequent traders" who only adjust their portfolios every $k \geq 2$ periods, and $1/k$ of them (total q/k mass) adjusts in each period. When they are not trading, their dividend income are reinvested at risk-free rate until their next investment decision. The remaining $1 - q$ mass of agents, which we can think of as investors who actively track flow information, or professional market-makers who

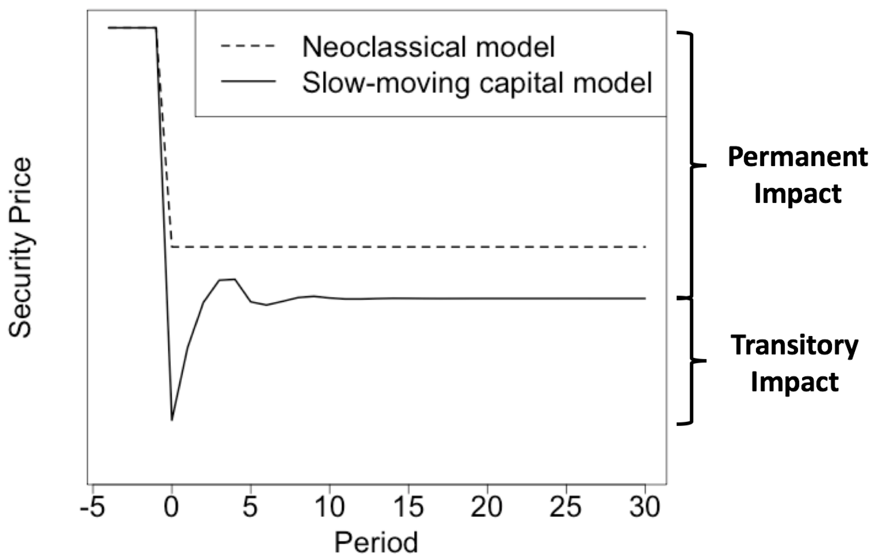

Figure 1. The price path following a large supply shock of 1 share of the asset at time 0. Parameters: $\delta = 0.05, Var(D_t) = Var(Z_t) = 0.1, \gamma = 1, r_f = 10\%$. In the slow-moving capital model, fraction $q = 0.8$ of traders participant only every $k = 5$ periods.

participate in the market at all times, are "frequent traders" and trade in all periods. Note that this model nests the neoclassical benchmark model with only frequent traders ($q = 0$).

There is exogenous, price-inelastic asset demand with level Z_t whose changes $f_t \equiv Z_t - Z_{t-1}$ correspond to factor-level fund flows we observe in the data. We assume mean-reverting flow dynamics:

$$f_{t+1} = -\delta Z_t + \epsilon_{t+1}^f, \quad \text{where } \delta \in (0,1), \quad \epsilon_t^f \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_f^2)$$

In the data, the mean-reversion parameter δ appears to be very small and statistically indistinguishable from zero, but we require that $\delta > 0$ in the model so that Z_t is stationary.

Equilibrium is characterized by agent portfolio optimizatino and market clearing. Duffie (2010) shows that, under 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 shares held by infrequent traders in period t. The positions held off the market show up because they affect the current aggregate security supply. Please consult the appendix in Duffie (2010) for solution details.

Relative to standard neoclassical models, due to slow-moving capital, flows generate excessive price impact that reverts over time. The mechanism is straight forward: when a supply shock hits, only a subset of agents – the $1-q$ frequent traders plus q/k mass of infrequent traders – are present to absorb the shock. For markets to clear, they demand a large initial price concession to bear the concentrated risk, and earn profits when later offloading their positions to other infrequent traders who arrive with a lag. As risk gets better shared among all agents, the excessive, transitory part of initial price impact reverts. Note that there is no trading between the exogenous demand submitters and optimizing agents in this reversion process. All trading happens among the optimizing agents themselves, and this is different from how reversion is generated in most neoclassical models, which we analyze below.

2.2 Differentiating from the standard neoclassical model

We now set $q = 0$ so all agents are always ready to absorb demand shocks in all periods. This is the assumption behind most, if not all, existing models that analyze the impact of exogenous demand shocks. Those models mostly differ on how the exogenous demand process is modeled, such as through price-inelastic noise traders (Campbell and Kyle (1993)), return-chasing agents (Barberis and Shleifer (2003)), or belief-extrapolating agents (Barberis, Greenwood, Jin, and Shleifer (2015)). They do not differ in the assumption about liquidity provider behavior.

Now that capital is perfectly mobile, Z_t is the only relevant state variable, and price is given by:

$$P_t = \lambda Z_t \tag{1}$$

Solution details are in the appendix. To illustrate this difference between these two models, in Figure 1, we simulate the price paths subsequent to a one-time large supply shock in both models. In the neoclassical model, price *immediately* adjusts to the post-shock equilibrium level with no reversion. In contrast, with slow-moving capital, asset price overshoots initially, and then reverts towards the new equilibrium level over time. While the neoclassical model only has permanent price impact, the slow-moving capital model features amplified volatility due to transitory movements[5]

One may be perplexed by the fact that many existing neoclassical models in the literature do generate return reversions. They do so via one of two ways. Note that expected excess return

[5]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.

decomposes into expected price changes and expected dividend yields. For all $h \geq 1$:

$$
\begin{aligned}
E_t(R_{t+h}) &\equiv E_t\left[P_{t+h} + \sum_{j=1}^{h} R_f^{h-j} D_{t+j}\right] - R_f^h P_t \\
&= \underbrace{E_t(P_{t+h} - P_t)}_{\text{price movement}} + \underbrace{\sum_{j=1}^{h} R_f^{h-j} \cdot \left[E_t(D_{t+j}) - r_f P_t\right]}_{\text{“dividend yield”}} \qquad (2)
\end{aligned}
$$

Some models, such as Barberis and Shleifer (2003), mainly generate reversion through the price movement term arising from mean-reverting demand shocks. In their model, the demand of "switchers" depend positively on recent price changes with exponentially decaying weight.[6] 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, switcher demand reverts back to normal levels and thus they sell. This gives rise to mean-reverting demand.[7]

The model in section 2.2 captures the same price-reversion generating mechanism as we assumed that flows mean-revert. However, if this is the true model, then after controlling for future flows, current flows cannot predict future price reversions. Formally:

Proposition 1.

For any horizon $h \geq 1$, in the neoclassical model, we have:

$$E(P_{t+h} - P_t | f_t, f_{t+1}, ..., f_{t+h}) = E(P_{t+h} - P_t | f_{t+1} + ... + f_{t+h}) \qquad (3)$$

This is generically not true in the slow-moving capital model.

The proof of (3) is immediate from (1):

$$P_{t+h} - P_t = \lambda(Z_{t+h} - Z_t) = \lambda \cdot (f_{t+1} + ... + f_{t+h})$$

In the slow-moving capital model, even after controlling for future flows $\{f_{t+1}, ..., f_{t+h}\}$, there

[6] See equations (9) and (10) in their paper.

[7] As another example, in Greenwood (2005), demand shock is permanent, but all securities liquidate at a predetermined horizon T upon which the demand shock effective reverts as it no longer matters for prices.

can be price reversion after the current flow shock:

$$\frac{\partial E(P_{t+h} - P_t | f_t, f_{t+1}, ..., f_{t+h})}{\partial f_t} < 0 \qquad \text{for all } f_{t+1}, ..., f_{t+h}$$

This is hard to prove analytically, but turns out true in all numerical solutions under a wide range of parameters.

The second way existing models generate return reversions (note, not price reversions) is through changing the dividend yields term in (2). Even if prices do not revert, after a supply shock, price levels are permanently lower and thus dividend yields are permanently higher. However, our later empirical analysis shows that almost all reversions come from price reversions rather than dividend yields changes.

3 Empirical Results

3.1 Data

Most equity mutual funds in the U.S. are benchmarked to capitalization or valuation based indices, as shown in the appendix.[8] Because fund managers cannot deviate from their target style too much, when investors purchase or redeem shares of mutual funds with different styles, that induces trades on stocks with corresponding styles. Lou (2012) documents that, in response to each dollar of out flow, mutual fund managers sell one dollar of stock positions in proportion to current allocations within the same quarter; the response is slightly delayed for in-flows as managers can hold cash and wait, but they still expand current positions by 60-80 cents for each dollar of in-flow. Thus, at a style level, fund managers are largely irrelevant, and the style allocation is mostly determined by end investor demand.

We use the standard CRSP survivorship-bias-free mutual fund database to get mutual fund returns and flow data. As is standard in the literature, we restrict to domestic equity funds[9] that on average dedicates 75% to 125% of its portfolio value to holdings stocks, and have no less than $1 million in total assets to avoid very small funds or incubation biases. We delete fund-month

[8]Morningstar, the largest mutual fund analysis company, classifies funds into a 3 by 3 style box exactly along the dimensions of size and value.

[9]CRSP Objective Code starting with "ED".

observations where fund return or asset under management (AUM) data is missing. Due to identifier availability limitations, we do analysis at the fund share class level, rather than fund level, because we only have identifiers (CRSP's "fundno") at the share class level before 1979.[10] After all data filters, including those described later, we end up with 14,054 unique funds and 456,261 unique fund-year observations.

For each year-end month t, for each fund i, we compute dollar net in-flows for that year:

$$flow_{i,t} = AUM_{i,t} - AUM_{i,t-1} \cdot \prod_{s=t-11}^{t} (1 + r_{i,s}^{\text{monthly}})$$

where $AUM_{i,t}$ is the total net asset under management as of the end of month t and $r_{i,t}^{\text{monthly}}$ is the post-fee fund return during month t. The way we calculate fund flows follows prior studies such as Coval and Stafford (2007),

To sort funds by their styles as defined by Fama-French factors, at each year end, we run time-series regressions of past 36 monthly fund returns against Fama-French 3 factor returns:

$$r_{i,\tau}^{\text{monthly}} = \alpha_i + b_{i,t} \cdot MKT_\tau + s_{i,t} \cdot SMB_\tau + h_{i,t} \cdot HML_\tau + \epsilon_{i,\tau} \qquad , \tau \in \{t-35, ..., t-1, t\}$$

Where MKT_t, SMB_t, HML_t are monthly Fama-French factor returns obtained from Ken French's website. We require all past 36 monthly returns to exist. From now on, we use the estimated SMB and HML loadings to measure fund styles.[11]

Reassuringly, the estimated loadings are strongly correlated with the self-reported fund objectives, suggesting we are truly measuring styles. Figure 2 plots the interquartile range of SMB and HML loadings or funds that have self-reported capitalization-based and valuation-based styles. Clearly, funds self-reported to have smaller capitalization styles have much higher SMB loadings. The correspondence is stronger for size exposure as the definition of capitalization-style is unambiguous. The relationship is weaker for value exposure because the industry definition of growth and value is slightly different from the Fama-French definition which uses book-to-market ratios

[10] CRSP provides a mapping from "fundno" to fund-level identifier "portno" but that is only available after 2001. One can also use Wermer's MFLINK to aggregate to fund level but that starts in 1979 Wermers (2000).

[11] Directly inferring style from the characteristics of fund holdings may be less noisy than our method(Daniel and Titman (1997)), but 13F portfolio filings data only goes back to 1980, while our sample starts in 1962. In unreported robustness tests, we used fund holdings to infer style after 1980 and found very similar results.

(Morningstar (2016)), but there is still a clear association. There is significant variation of styles across funds. Throughout the sample, 55% and 51% of the fund-year observations have significant t-statistics at the 5% level for size and value loadings, respectively.

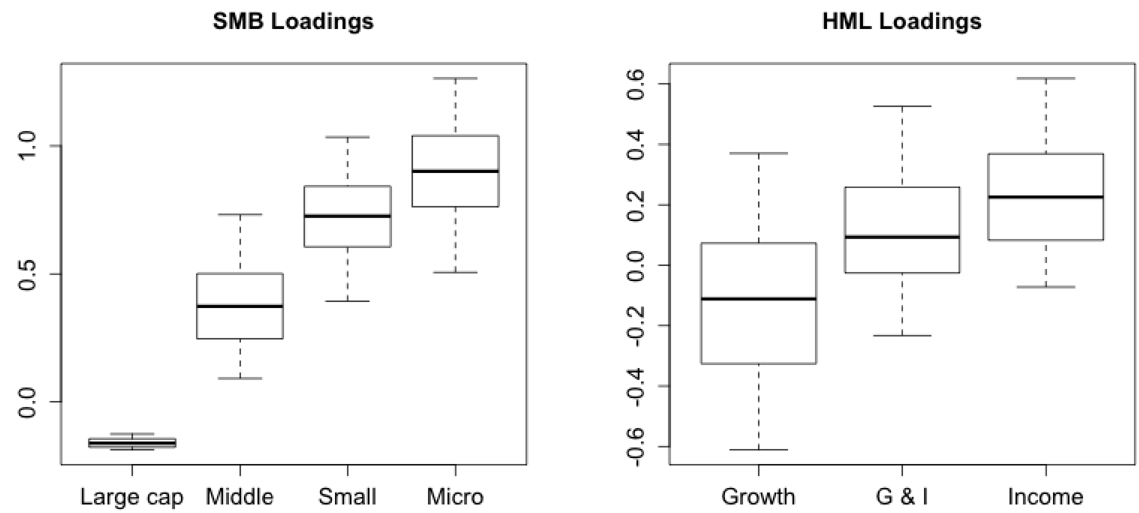

Figure 2. Distribution of style loadings plotted against self-reported fund objectives. The box represent 25% to 75% percentiles, while the whiskers are 5% and 95% percentiles.

After applying all data filters mentioned above, we end up capturing close to a little more than half of the entire equity mutual fund industry. The CRSP database starts in 1962 but as we need 3 years of data to estimate fund style, our sample starts in 1965 and ends in 2015. Summary statistics are collected in Table 1.

Period	Fund-year	Unique	AUM (MM USD)		AUM Percentile			Annual	Flow/
	Obs	Funds	Total	Mean	25%	50%	75%	Return	AUM
1965-1970	687	136	27,885	244	16	79	283	3.9%	4.3%
1971-1975	987	247	36,513	185	17	51	187	1.0%	-4.7%
1976-1980	1,173	247	31,956	136	16	41	125	17.8%	-11.0%
1981-1985	1,274	287	54,623	214	38	81	233	14.1%	3.5%
1986-1990	1,969	513	143,427	364	51	125	364	9.5%	-0.7%
1991-1995	3,466	880	382,692	552	48	150	463	16.2%	12.3%
1996-2000	9,002	2,935	1,701,773	945	39	150	566	12.8%	4.9%
2001-2005	23,196	6,615	2,643,091	570	16	67	279	1.8%	1.0%
2006-2010	31,505	8,457	3,502,951	556	11	55	253	2.9%	-1.9%
2011-2015	38,728	9,719	5,161,854	666	14	71	325	10.4%	-1.4%

Table 1. Summary of domestic equity mutual funds used in this study. The Total AUM is the average over the corresponding period. Fund return and flows are AUM weighted.

3.2 Factor structure in mutual fund flows

In order to detect factor structure, in each year, we double sort funds by their SMB and HML loadings into 5 by 5 portfolios and perform all subsequent work at the portfolio level. This is similar to how Fama and French (1996) double sorts stocks, except we are sorting on funds and looking at flows rather than returns. For each portfolio j, we compute the sum of net fund flows. Because mutual fund AUM and overall stock market capitalization grew significantly over this 51 year period, we divide flows by the 12-month ago portfolio AUM to get stationarity:

$$f_{j,t} = \frac{\sum_{i \in \text{ portfolio} j} flow_{i,t}}{\sum_{i \in \text{ portfolio } j} AUM_{i,t-12}} \tag{4}$$

Principal component analysis on flows of the 25 portfolios reveals a clear factor structure. As plotted in Figure 3, the first three principal components explain 30%, 16%, and 11% of variance, respectively. If flows are random, then each of the 25 PCs should explain 4% of variance in population. To correct for small sample bias, we also compare the numbers against that obtained from simulated, normally distributed random flows with the same sample length. The first three PCs still appear large. Interestingly, when comparing to this small sample benchmark, only the first three PCs stand out as abnormally large.

More interestingly, the first 3 PCs appear to be linear combinations of Fama-French 3 factors. To see this, we plot portfolio loadings on the first 3 PCs in Figure 4. All 25 portfolios have roughly equal loading on the first PC which is similar to the market factor, and the next two PCs appear to be linear combinations of the size and value factors. The PC2 portfolio is long small-value stocks and short large-growth stocks, while the PC3 portfolio is long small-growth and short large-value.

One may worry that, because fund flows chase returns, this flow factor structure merely inherits the factor structure well documented in returns. This is not the case. In appendix C.1, we show that the factor structure is roughly unchanged, both in terms of magnitude and in terms of how the factors resemble Fama-French factors, after regressing out lagged returns from fund flows.

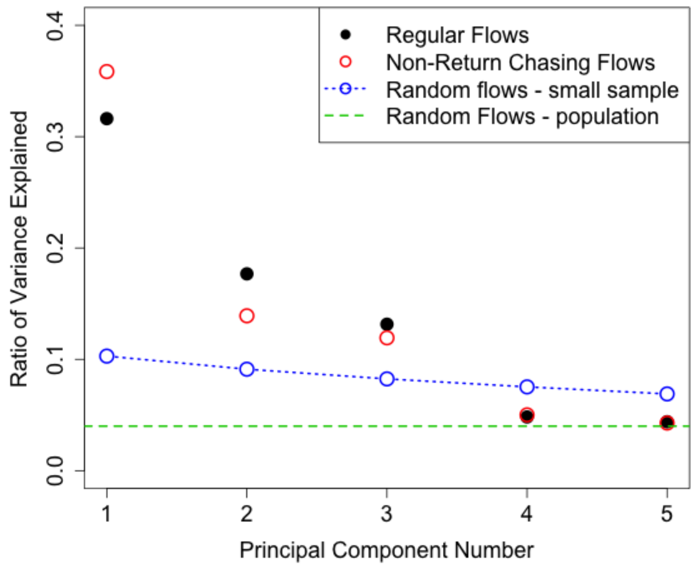

Figure 3. Ratio of variance explained by first principal components of 5 by 5 portfolio flows. Dashed line is what the in-population ratio should be ($1/25 = 4\%$) if flows are random.

3.3 Factor-level flows are large and persistent

To inspect flows at the factor level, we follow the Fama-French procedure to construct the fund flow analog of SMB and HML. In each year, we double sort funds into 2 by 3 equal weighted bins using SMB and HML loadings, and define the flow factors in the obvious fashion:

$$fSMB_t = \frac{f_{SH,t} + f_{SM,t} + f_{SL,t}}{3} - \frac{f_{BH,t} + f_{BM,t} + f_{BL,t}}{3}$$
$$fHML_t = \frac{f_{SH,t} + f_{BH,t}}{2} - \frac{f_{SL,t} + f_{BL,t}}{2}$$

where $f_{SH,t} = \frac{\sum_{i \in \text{(small cap, high book/market)}} f_{i,t}}{\sum_{i \in \text{(small cap, high book/market)}} AUM_{i,t-12}}$, etc.

The flow factors are large and persistent, making it plausible that they can create large price impacts. The left panel of figure 5 plots the time series of flow factors against corresponding Fama-French return factors, and show that with $\pm 20\%$ flows are very common in a given year. To be clear, a 20% in flow in fSMB, roughly implies that investors increase small cap holdings by 10% and decrease large cap holdings by 10%. As shown in the right panel, the flows are also persistent, with statistically significant autocorrelations out to almost 2 years for fSMB and 1 year for fHML. There is no statistically significant sign of flow reversion for fSMB and only some slight reversion of

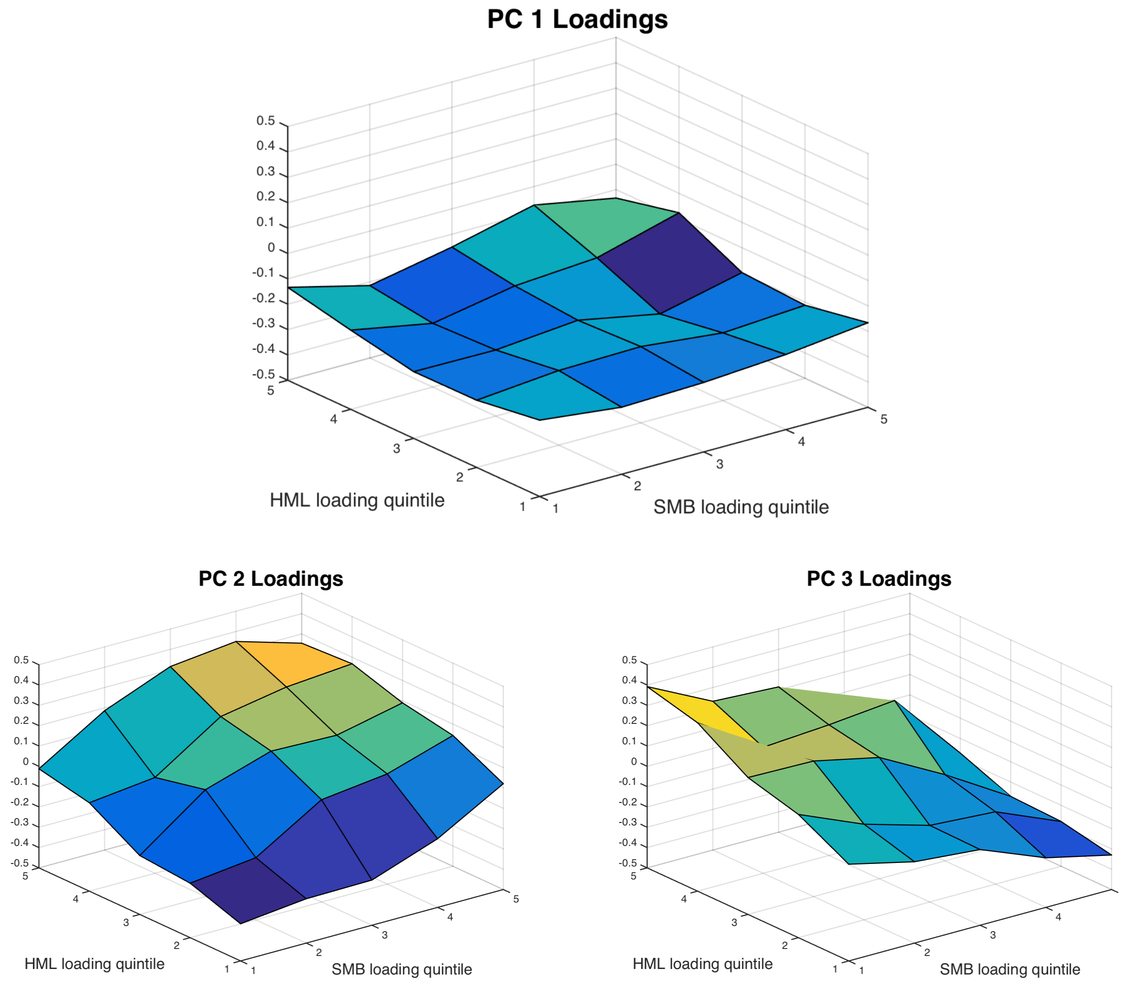

Figure 4. Loadings of the first 3 principal components in the 25 size - book/market mutual fund portfolio flows.

fHML at 5-6 years horizon.

To get a sense of the large flow magnitude, the out flow from SMB is \$41billion in 1997, \$133 billion in 1998, and \$158 billion in 1999. That is \$332 billion dollars of selling in three years. The average total U.S. equity capitalization at the time was \$13.7 trillion, so the net selling amounts to 2.4% of the market. The continued fHML in-flows during 2001 - 2005 add up to \$640 billion dollars of net buying, which is around 4.1% of the entire market capitalization at that time. Consistent with large flows generating price impact, fSMB and fHML are positively correlated with return factors with 33% and 18% correlations, respectively.

We only measure style-level flows in mutual fund share reallocations, but it is worth bearing in mind that the true underlying flow may be a few times larger. Using data from Federal Reserve Flow

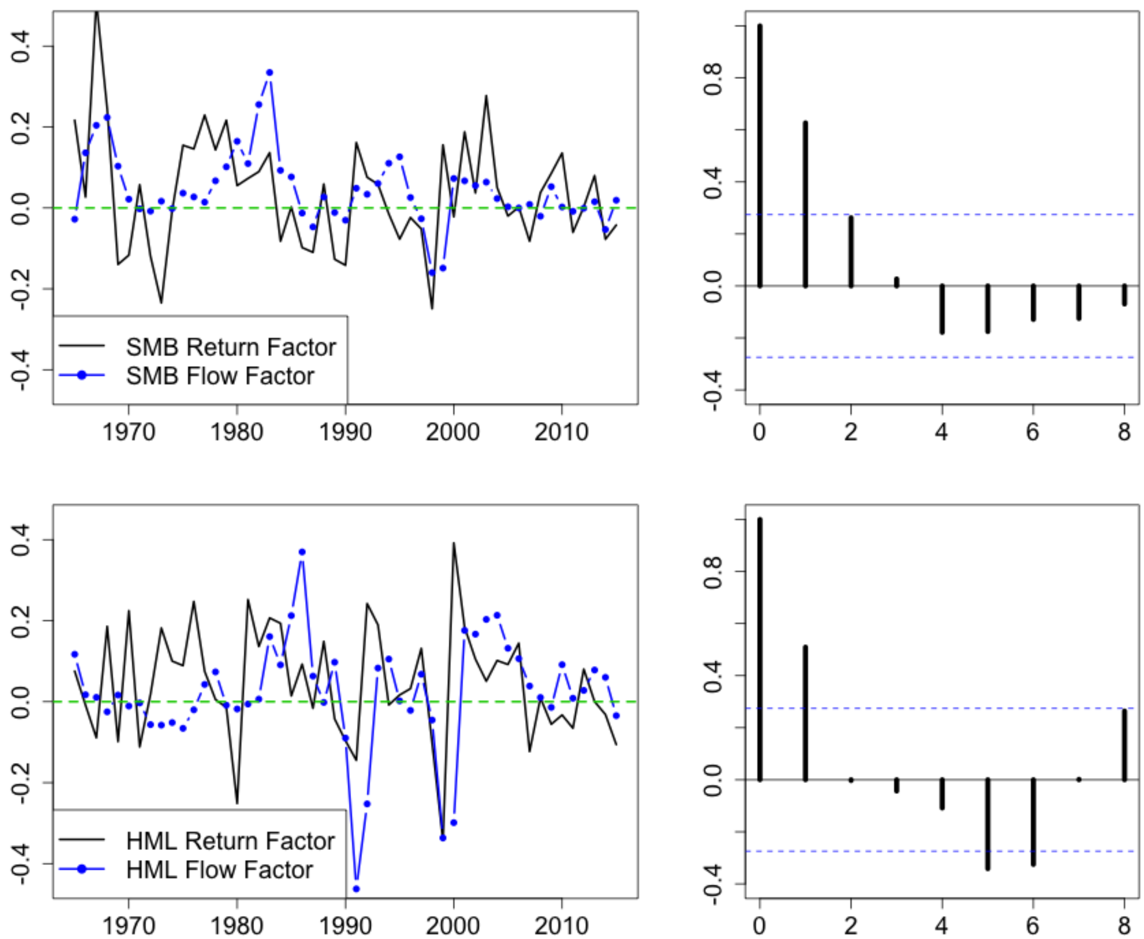

Figure 5. Left: SMB and HML flow factors plotted against return factors. Right: Autocorrelation function of flow factors.

of Funds, Figure 6 breaks down the fraction of U.S. equities held by different sectors and compare against the mutual funds we capture in our study. First of all, due to our data filters and coverage limitations of the CRSP mutual fund database, we only capture half of the mutual fund sector by AUM, but it is very likely that similar style-level flows happen to other mutual funds. Therefore the true mutual fund style flow can be twice as large as what we measured. Further, if mutual fund flows reflect time-varying retail investor views at the style level, then it is plausible that such investors are conducting similar style shifts in their other direct or indirect portfolio holdings, in which case the true flow may be 3-5 times larger than we measure. Throughout the paper, we only look at flows in the mutual funds in our data, but when estimating demand elasticities in section 5, we discuss how different assumptions impact the conclusion.

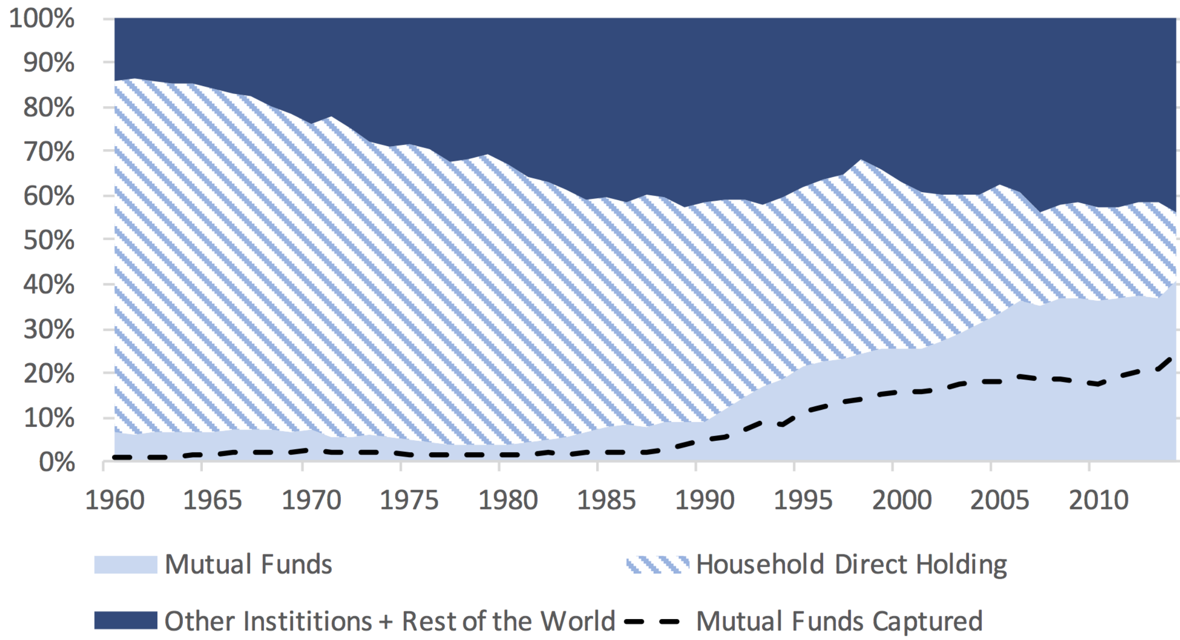

Figure 6. Percentage of U.S. publicly traded equity held by different sectors. The black dotted line represents the hold of all mutual funds that we study in the paper.

3.4 Fund flows are not informed about cash flows

If mutual fund investors have superior information about future stock returns or cash flows, then their trades should create price impact according to standard asymmetric information models such as Kyle (1985). In this section we show that fund flows do not seem to carry any forecasting power about cash flows. Combined with the fact that flow factors negatively forecast return factors, and Frazzini and Lamont (2008)'s evidence that retail reallocation across mutual funds is wealth destroying, we argue it is safe to assume that mutual fund flows do not carry superior information, and the slow-moving capital model characterization of fund flows as exogenous supply shocks, likely due to preference or belief changes, is appropriate.

To proxy for style level cash flows, we compute return on equity (ROE) and return on asset (ROA) for the 2 by 3 size-value portfolios and construct SMB and HML cash flow factors in the obvious way. We then run forecasting regressions to see whether factor fund flows predict factor cash flows. Because ROE and ROA are highly persistent measures, we add their own lags as controls.

The regression results in Table 2 show that there is no evidence of fund flows positively predicting cash flows. The reverse is not true: past cash flows do have marginal power in predicting future

	Dependent Variable = cash flow$_t$							
	Cash Flow = ROE				Cash Flow = ROA			
	SMB	HML	SMB	HML	SMB	HML	SMB	HML
Intercept	−0.01	−0.03*	−0.00	−0.03	−0.01**	−0.01*	−0.01**	−0.01*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
fund flow$_{t-1}$	0.02	−0.07*	0.07	−0.04	0.00	−0.02	0.01	−0.00
	(0.03)	(0.03)	(0.04)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)
fund flow$_{t-2}$			−0.08	−0.03			−0.00	−0.03*
			(0.05)	(0.05)			(0.01)	(0.02)
fund flow$_{t-3}$			−0.05	−0.05			−0.01	0.00
			(0.04)	(0.04)			(0.01)	(0.01)
cash flow$_{t-1}$	0.84***	0.65***	0.66***	0.60***	0.66***	0.75***	0.61***	0.72***
	(0.08)	(0.12)	(0.15)	(0.16)	(0.11)	(0.10)	(0.15)	(0.15)
cash flow$_{t-2}$			−0.05	−0.18			0.19	0.11
			(0.17)	(0.18)			(0.18)	(0.17)
cash flow$_{t-3}$			0.34*	0.28			−0.22	−0.10
			(0.14)	(0.15)			(0.15)	(0.14)
Adj. R^2	0.74	0.39	0.78	0.41	0.42	0.56	0.44	0.62
N	49	49	47	47	49	49	47	47

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2. Forecasting regression of annual SMB and HML factor-level cash flow measures using past fund flows and cash flows.

fund flows as shown in the appendix (Table 7).

3.5 Factor flows predict factor price reversals

If the slow-moving capital model in section 2 is the right model, then factor flows must predict factor price reversals. To be consistent with prior literature on return predictability, we use future factor total returns as the left-hand side variable, but show in the appendix that almost all predictability comes from predicting price reversals rather than dividend returns. Recall from section 2.2 that this is the key difference from neoclassical models: in response to demand shocks that do not revert, which is approximately the case for mutual fund flows, neoclassical models generate return reversions only through changing dividend yields while the slow-moving capital model also generates price reversals.

he reversion evidence is already visually clear in a scatter plot of annualized 5 year factor returns against lagged factor flows (Figure 7). To do this more formally, for each year end month t and for each factor $F \in \{SMB, HML\}$, we regress future 5 year annualized factor returns on lagged one

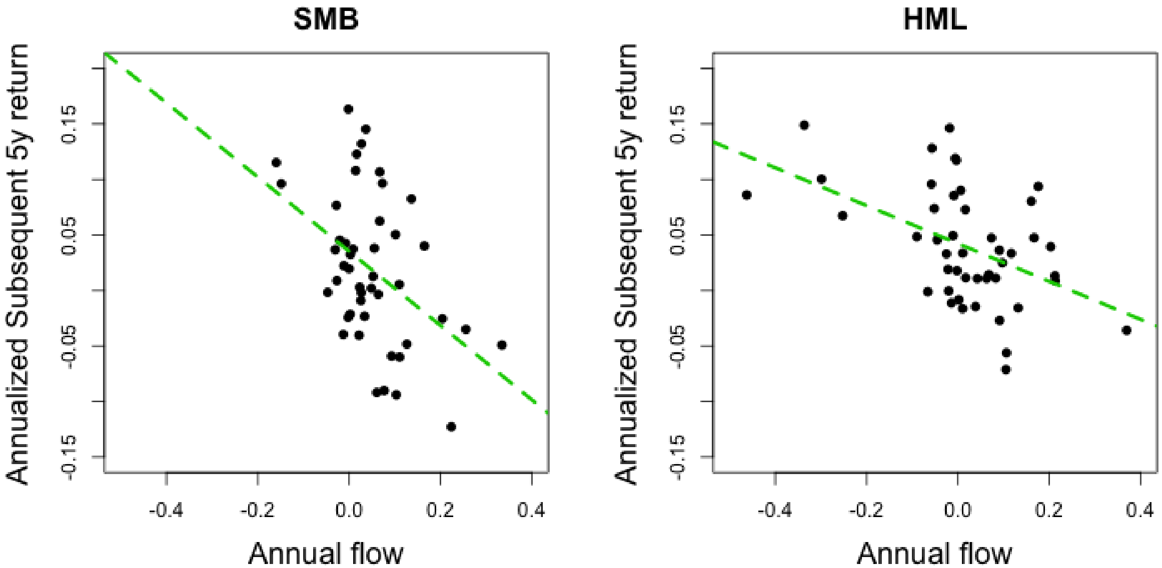

Figure 7. Annualized five year forward factor returns plotted against one-year factor flows. Green dashed line is OLS fit.

year factor flows:

$$\frac{\text{ret}^F_{t\to t+60}}{5} = a^F + b^F \cdot \text{flow}^F_{t-12\to t} + \epsilon^F_{t\to t+60} \tag{5}$$

Regression results are reported in columns 2 and 3 in table 3 and standard errors are adjusted following Hodrick (1992). The forecasting relationship is statistically and also economically significant: in response to one standard deviation flow shock (0.087 for SMB and 0.138 for HML), future average annual return is lowered by 2.9% and 2.5% for SMB and HML, respectively. The R^2 is around 20%, and simulation under a null of no predictability gives p values less than 0.001 of obtaining such high R^2. Apart from Greenwood and Hanson (2012) and Cohen, Polk, and Vuolteenaho (2003), we are not aware of any other paper finding such high factor-level return predictability.

	Main Regression		Using Flow Residuals		Controlling for Future Flows	
	SMB	HML	SMB	HML	SMB	HML
Intercept	0.038**	0.045**	0.021	0.041**	0.014	0.040*
	(0.018)	(0.020)	(0.018)	(0.021)	(0.022)	(0.022)
$\text{flow}_{t-12\to t}$	−0.338***	−0.179**	−0.310**	−0.162	−0.406***	−0.180*
	(0.123)	(0.071)	(0.143)	(0.100)	(0.122)	(0.071)
$\text{flow}_{t\to t+60}$					0.129**	0.048
					(0.044)	(0.040)
Adjusted R2	0.172	0.213	0.068	0.076	0.455	0.346
N	46	46	43	43	46	46

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 3. Forecasting annualized 5 year factor returns using annual factor flows. Column 4 and 5 uses factor flow residuals as the right hand side. The last two columns controls for future flows in the regression. Standard errors are calculated following Hodrick (1992).

4 Alternative Explanations

We now consider two main alternative explanations for our findings. We identify what distinguishes those stories from the slow-moving capital model and test the predictions in the data.

4.1 The return chasing hypothesis

One alternative hypothesis is that flows do not have any causal impact on prices, but are merely chasing past returns as documented in the prior literature (e.g. Coval and Stafford (2007)), and appear to predict future returns due to exogenous mean-reversion of factor returns. The following toy model makes this view concrete. Again consider a single asset which is interpreted to be the SMB or HML portfolio in our paper. Suppose asset price follows an exogenous mean-reversion process:

$$P_{t+1} = \rho \cdot P_t + \epsilon_{t+1}^P$$
$$\text{where } 0 < \rho < 1, \quad \epsilon_{t+1}^P \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\epsilon^P}^2)$$

Recall that dividends are independently distributed as $D_t \sim \mathcal{N}(0, \sigma_D^2)$. Suppose measured flows chase price movements, and because of infrequent measurement, show up as being positively dependent on concurrent price changes:

$$f_t = \beta \cdot (P_t - P_{t-1}) + \epsilon_t^f, \quad \beta > 0, \ \epsilon_t^f \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\epsilon^f}^2)$$

Where ϵ_t^f is independent from every other variable. It captures both the non-return chasing part of flows and possible classical measurement errors. Under these assumptions, contemporaneous flows will forecast future price changes. To see this, note that for any $h \geq 1$, we have:

$$
\begin{aligned}
Cov(f_t, P_{t+h} - P_t) &= \beta Cov\left(P_t - P_{t-1}, P_{t+h} - P_t\right) \\
&= \beta Cov\left(-(1-\rho)P_{t-1} + \epsilon_t^P, \quad \left[\sum_{j=1}^{h+1} \rho^{j-1}\epsilon_{t+h-j+1}^P + \rho^{h+1}P_{t-1}\right] - \left[\rho P_{t-1} + \epsilon_t^P\right]\right) \\
&\overset{Var(P_t)=\frac{Var(\epsilon_t^P)}{1-\rho^2}}{=} \beta \cdot \left(\rho(1-\rho)(1-\rho^h)Var(P_{t-1}) - (1-\rho^h)Var(\epsilon_t^P)\right) \\
&= \beta \cdot \left[\frac{\rho(1-\rho^h)}{1+\rho} - (1-\rho^h)\right] \cdot Var(\epsilon_t^P) \\
&= -\frac{\beta(1-\rho^h)}{1-\rho} \cdot Var(\epsilon_t^P) < 0
\end{aligned}
$$

If this view is true, then if we take out the return-chasing part of flows, the residual flow should have no predictive power over future returns. In other words:

$$
Cov(\epsilon_t^f, P_{t+h} - P_t) = 0 \quad \text{for all } h \geq 1
$$

We test this view empirically by first regressing fund flows on current and lagged 4 years of fund returns (Table 4), and then reconstructing flow factors using the flow residuals. We then use the resulting, residualized flow factors to predict future factor returns. The results are shown in column 4 and 5 in Table 3. Although statistical significance and coefficient value decrease to some extent, the forecasting power of flows is far from explained away. The coefficient for SMB is still significant at 5% level, and while the HML coefficient loses statistical significance, the coefficient size is only attenuated by around 25%. We verify in the appendix that predicting price returns, rather than total returns, yield the same result. Thus we conclude that mechanical return-chasing does not explain our findings.

	Fund Flow_{t-1}
(Intercept)	-0.10^{***}
	(0.02)
Fund Return_t	0.30^{***}
	(0.09)
Fund Return_{t-1}	0.26^{***}
	(0.05)
Fund Return_{t-2}	0.23^{***}
	(0.04)
Fund Return_{t-3}	0.14^{***}
	(0.02)
Fund Return_{t-4}	0.14^{**}
	(0.04)
Num. obs.	77637
Adj. R^2	0.04

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4. Panel regression of annual fund flows on current and lagged fund returns. Standard errors are clustered by year and fund.

4.2 Neoclassical models with heterogeneous agents

As explained in section 2.2, the return predictability from flows can still be consistent with neoclassical models with heterogenous agents and mean-reverting flows. However, as Proposition 1 shows, in a neoclassical model where liquidity providers are not slow to react, past flows should have no predictive power after we control for future flows in the regression. Specifically, we run the following modified regression:

$$\frac{\text{ret}^F_{t\to t+60}}{5} = a^F + b^F \cdot \text{flow}^F_{t-12\to t} + \text{flow}^F_{t\to t+60} + \epsilon^F_{t\to t+60}$$

Results are shown in the last two columns of Table 3. The predictive power of lagged flows is not diminished at all, if not enhanced, after controlling for future flows. The appendix reruns the regression using only price returns and finds the same qualitative result. Thus we conclude the neoclassical view is inconsistent with evidence.

This should be unsurprising as there is little, if any, evidence of mean-reversion in flows (right panel of Figure 5). There is some reversion evidence in HML flow at 5-6 year lag, but the HML reversion mostly happen within 1-2 years, so the reverting flow cannot be the cause of the reverting returns.

5 Structural estimation

In order to isolate the component of factor returns caused by fund flows, we fit a bivariate structural vector autoregressive model (SVAR) with 5 lags to the annual flows and returns of each factor. For each factor $F \in \{SMB, HML\}$, let $y_t^F = (r_t^F, f_t^F)'$ denote the demeaned vector of log returns and flows. We omit index F later for notational simplicity. The SVAR is specified as:

$$\begin{pmatrix} r_t \\ f_t \end{pmatrix} = \begin{pmatrix} \lambda f_t \\ \beta r_t \end{pmatrix} + \begin{pmatrix} B_{1,rr} & B_{1,rf} \\ B_{1,fr} & B_{1,ff} \end{pmatrix} \begin{pmatrix} r_{t-1} \\ f_{t-1} \end{pmatrix} + ... + \begin{pmatrix} B_{5,rr} & B_{5,rf} \\ B_{5,fr} & B_{5,ff} \end{pmatrix} \begin{pmatrix} r_{t-5} \\ f_{t-5} \end{pmatrix} + \begin{pmatrix} u_t^r \\ u_t^f \end{pmatrix} \tag{6}$$

where λ is the price impact coefficient, β captures the fact that flows can chase returns within the same year, and $u_t = (u_t^r, u_t^f)'$ are the structural innovations. Because there is contemporaneous feedback effect between flows and returns, we need an additional assumption to achieve identification. Following Blanchard and Quah (1989), we impose that the long-run impact of flows on cumulative price changes is zero. We argue this is a plausible assumption because the liquidity providing sector has, throughout the sample period, approximately zero net position on both SMB and HML, so absorbing additional factor flows does not imply large risk increases relative to their existing portfolios. This is also an empirically motivated assumption as price impacts in the data do appear to fully revert. However, we do admit this is a strong assumption and could potentially bias the results.

We estimate the system by first estimating the reduced form VAR, and then using a Cholesky decomposition to get the structural parameters. The standard errors are derived by simulating the reduced form VAR through bootstrapping the error term and then applying the whole estimation procedure. The bootstrap sample size is 10,000.

5.1 Separating the flow-induced component of returns

Figure 8 plots the cumulative responses to a one-standard deviation flow shock out to 10 years. A one standard deviation flow shock leads to around 5-8% contemporaneous factor return movement, which subsequently reverses. The initial reversal is slower for SMB, consistent with the fact that

SMB flows are more persistent in the short run of one to two years (Figure 5). For either factor, the reversal is fully achieved by year 6. After a flow shock, more flows in the same direction continue out to 4 years for SMB and 2 -3 years for HML.

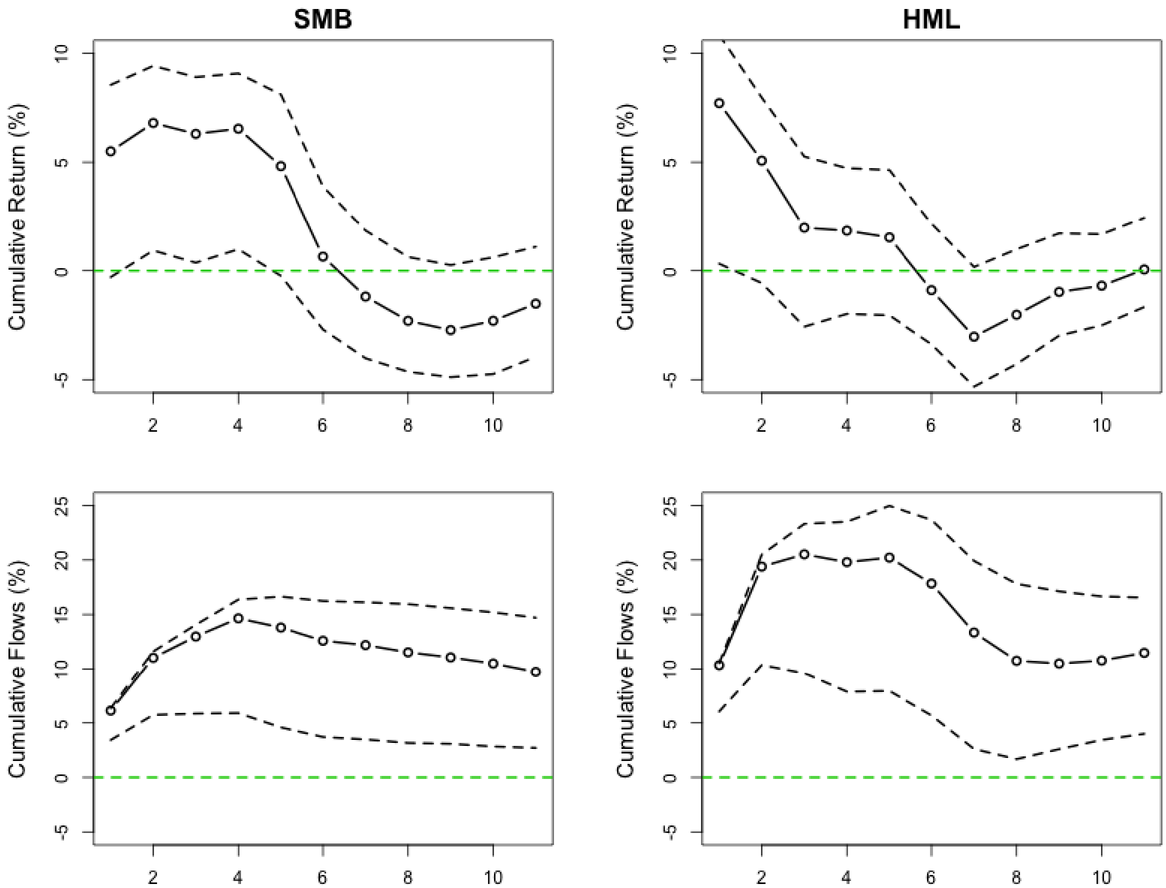

Figure 8. Impulse response function in the SVAR(5) system to flow shocks with 90% confidence intervals.

In the counterfactual world where all flows are zero, (6) reduces to a univariate autoregression:

$$r_t = B_{1,rr}r_{t-1} + ... + B_{5,rr}r_{t-5} + u_t^r \tag{7}$$

We simulate (7) while preserving the parameter point estimates in (6). The distribution of factor return standard deviation without flows is plotted against that with flows, derived by simulating the full SVAR, in Figure 9. Using the mean of the simulations, we estimate that in the case without flows, SMB and HML return volatility goes from 13.1% to 10.9% and from 12.5% to 10.6%, respectively. This means a reduction of variance by 31.2% for SMB and 28.1% for HML.

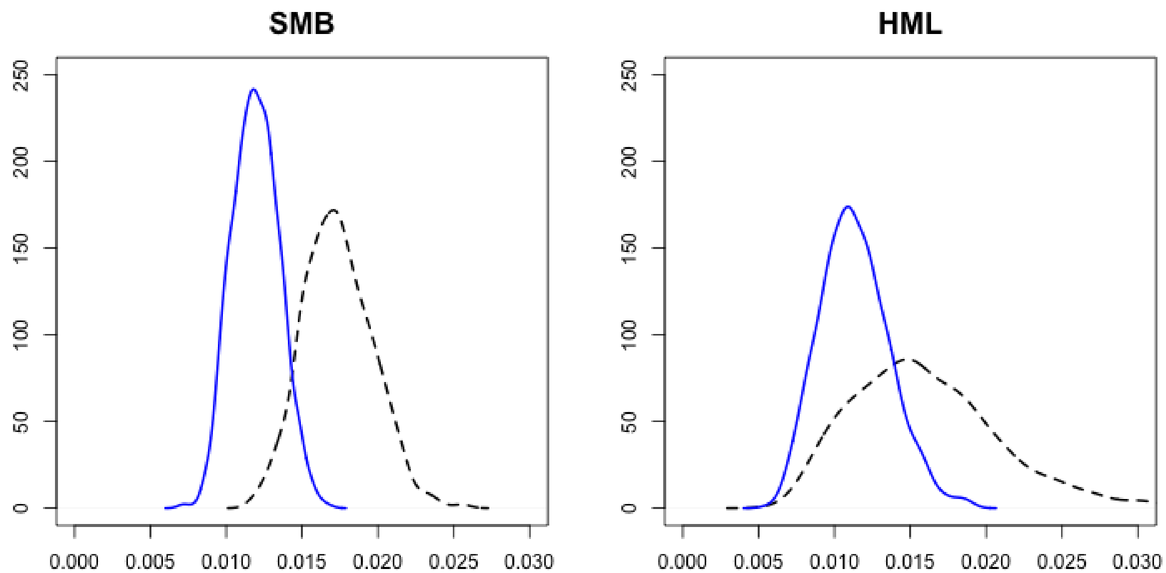

Figure 9. Kernel density of simulated annual factor return variance in the full SVAR (6) or the counterfactual model (7) without flows. Simulation sample size is 1,000.

5.2 Comparing demand elasticity and reversion speed with existing literature

We now investigate how the price impact coefficients in the structural model compare with that found in previous papers looking at uninformed demand shocks. Figure 10 plots the distribution of the bootstrapped λ estimates. The point estimates for SMB and HML are surprisingly similar, at 0.71 an 0.7, respectively. There is substantial uncertainty around the point estimates, and the p-values under the null of $\lambda = 0$ are 0.089 and 0.059, respectively.

What demand elasticity does this imply? That depends on whether we assume the flows we measure only represent trading through the mutual funds we measure, which is arguably a lower bound, or if the flows also proxy for similar trading behavior of the entire household sector, which arguably is an upper bound. In the former case, since mutual fund holdings is on average 16% of the market over the sample period, this implies an elasticity of $1/(0.7/16\%) = 0.22$. In the latter case, because household holdings is on average 67% of the total market capitalization over the sample period, this implies an elasticity of $1/(0.7/67\%) = 0.96$.

This estimate falls in a similar range with the index inclusion studies. Using a large sample of additions and deletions between Russell 1000 and Russell 2000 indices, Chang et al. (2014) found an elasticity of 0.39 - 1.46, which is only slightly higher than the high range of our estimate. Using

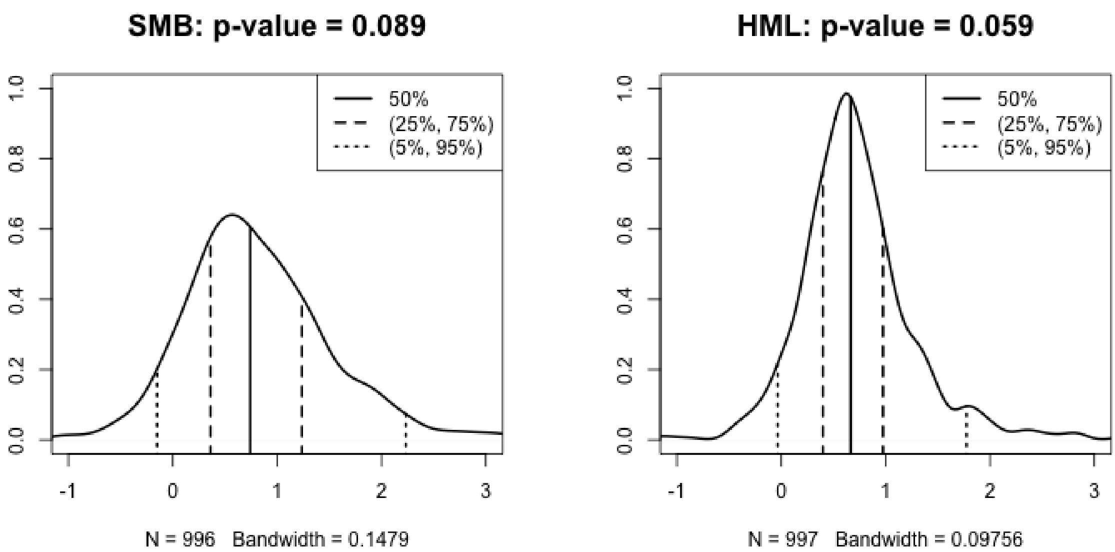

Figure 10. Kernel density distribution of bootstrapped price impact coefficient λ in SVAR(5).

a smaller set of S&P 500 index changes, Shleifer (1986) found an elasticity of around 1.

The price reversion takes years in our data, and this is consistent with the two prior papers looking at mutual fund flow induced price pressures. Coval and Stafford (2007) look at how fire sales and purchases of mutual funds, defined as flows above 90% or below 10% percentiles, impact the price of their stock holdings. Their figure 2 shows that prices continue to revert at least throughout the first 19 months after fire-sale events with no sign of stopping at 18 months, which is where their post-event window ends. Similarly, figure 2 in Edmans, Goldstein, and Jiang (2012) shows that reversion takes on average 2 years to complete.

6 Conclusion

Using mutual fund flows to measure retail investor reallocation of capital across stock styles, we find strong factor structure in their order flows along size and value directions. The style-level reallocations are i) large, frequently amounting to a few percentage points of the entire market capitalization outstanding, ii) persistent, with significant autocorrelation out to 1 year, and iii) apparently uninformed about future cash flows. Consistent with the hypothesis that other investors are slow to react and provide liquidity, fund flows create significant contemporaneous price impact that reverses in subsequent years, leading to strong factor return predictability. We found that a one standard deviation higher (or lower) factor-level fund flow leads to 2-3% lower (or higher) annual factor expected return in the subsequent 5 years.

We show that slow-moving liquidity provision is key to the reversion patterns. Standard rational or behavioral models assume liquidity providers react to flow shocks immediately, so they only generate price reversion if flows are mean-reverting. However, flows do not revert in the data. Slow-moving liquidity provision also explains why flows can generate the large flow-induced price movements we observe, which standard models have trouble accounting for without assuming risk aversion coefficients much larger than typically measured by microeconomic studies.

We fit a structural autoregressive model to capture joint dynamics of flows and returns. Reassuring, the estimated price impact coefficients and reversion speed are comparable to those found by the existing literature. Using the structural model, we estimate that fund flows explain around 30% of SMB and HML factor price variations. That is, if flows were absent, return factor variance would be 30% lower. In summary, this paper provides evidence that uninformed demand can generate large, persistent price movements at the broad-market level, and presents evidence that slow-moving liquidity provision plays an important role in amplifying and prolonging price impacts.

APPENDIX

A Solving the neoclassical model

We conjecture that the price is linear in the state variable:

$$P_t = \lambda Z_t \tag{8}$$

So positive exogenous demand raises prices. Merely for the purpose of analytical simplicity, we assume agents do myopic portfolio optimization. This is without loss of generality as Appendix C of Kozak, Nagel, and Santosh (2017) shows that the same pricing function is obtained when agents do long-horizon optimization[12]. We solve for agent's optimal demand curve $x_t^*(P_t)$:

$$\begin{aligned}
R_{t+1} &\equiv P_{t+1} + D_{t+1} - R_f P_t \\
&= \lambda Z_{t+1} + D_{t+1} - R_f P_t \\
E_t(R_{t+1}) &= \lambda E_t(Z_{t+1}) - R_f P_t \\
&= \lambda(1-\delta)Z_t - R_f P_t \\
Var_t(R_{t+1}) &= Var_t(P_{t+1} + D_{t+1}) \\
&= \lambda^2 \sigma_f^2 + \sigma_D^2 \\
\Rightarrow x_t^*(P_t) &= \frac{E_t(R_{t+1})}{\gamma Var_t(R_{t+1})} = \frac{\lambda(1-\delta)Z_t - R_f P_t}{\gamma(\lambda^2\sigma_f^2 + \sigma_D^2)}
\end{aligned}$$

Market clearing verifies the conjectured price form in (8):

$$\begin{aligned}
x_t^*(P_t) + Z_t &= 0 \\
\Rightarrow P_t &= \underbrace{\left(\frac{\lambda(1-\delta) + \gamma(\lambda^2\sigma_f^2 + \sigma_D^2)}{R_f}\right)}_{\overset{\text{need to}}{=}\lambda} \cdot Z_t
\end{aligned}$$

λ is a root of a quadratic equation which has real solutions if flow and dividend shocks are not

[12]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.

too volatile, or if risk aversion is not too high[13].

B Prevalence of mutual fund benchmarking to style indices

Most mutual funds are benchmarked to indices that are defined using rules on the market capitalization or growth/value characteristics of stocks. Petajisto (2013) has data on which benchmarks are followed by a large sample of U.S. mutual funds, and that is summarized in Table **??**. Most mutual funds (97.8% of AUM), except those following Russell 3000 and Wilshire 4500/Wilshire 5000, are benchmarked to some index with styles.

Index Name	Index Style	AUM in 2000 (MM USD)	AUM Ratio
S&P 500	Large Cap	1,355,615	58.8%
Russell 1000 Growth	Large/Mid Cap, Growth	241,215	10.5%
Russell 1000 Value	Large/Mid Cap, Value	141,817	6.1%
Russell Mid Cap Growth	Mid Cap, Growth	94,553	4.1%
Russell 3000 Growth	All Cap, Growth	86,999	3.8%
Russell 2000 Growth	Small Cap, Growth	73,990	3.2%
Russell 2000	Small Cap	61,558	2.7%
Russell 1000	Large/Mid Cap	48,881	2.1%
S&P 500/Barra Growth	Large Cap, Growth	42,626	1.8%
S&P Midcap 400	Mid Cap	27,637	1.2%
Wilshire 5000	All Cap	23,814	1.0%
Russell 3000	All Cap	21,186	0.9%
Russell 3000 Value	All Cap, Value	20,992	0.9%
Russell 2000 Value	Small Cap, Value	19,248	0.8%
Russell Mid Cap	Mid Cap	11,502	0.5%
S&P Small Cap 600	Small Cap	10,796	0.5%
Russell Mid Cap Value	Mid Cap, Value	9,661	0.4%
S&P 500/Barra Value	Large Cap, Value	9,090	0.4%
Wilshire 4500	All Cap	5,936	0.3%

Table 5. Benchmarking of U.S. equity mutual funds. Data downloaded from Petajisto's website.

[13] Specifically, the condition for having real solutions is:

$$\delta + r_f \geq 2\gamma\sigma_f\sigma_D$$

When the above condition is satisfied with an inequality, both roots are positive, so there are two solutions. We pick the low volatility one following prior literature (Bogousslavsky (2016)).

C Additional empirical results

C.1 PCA on non-return chasing flows

As explained in section 3.2, one may worry that the factor structure in fund flows mechanically arise from factor structure in returns and the fact that flows chase returns. To get around this concern, we run panel regression of annual fund flows on past 4 annual returns, and redo the PCA exercise using residuals. The factor structure is still clearly present. Figure **??** shows that the variance explained by first 3 PCs are roughly unchanged by taking out the return-chasing part of flows. Table **??** shows the loadings of the first 3 principal components before and after regressing out past returns. Taking out the return-chasing part of flows has almost no effect on the PC loadings.

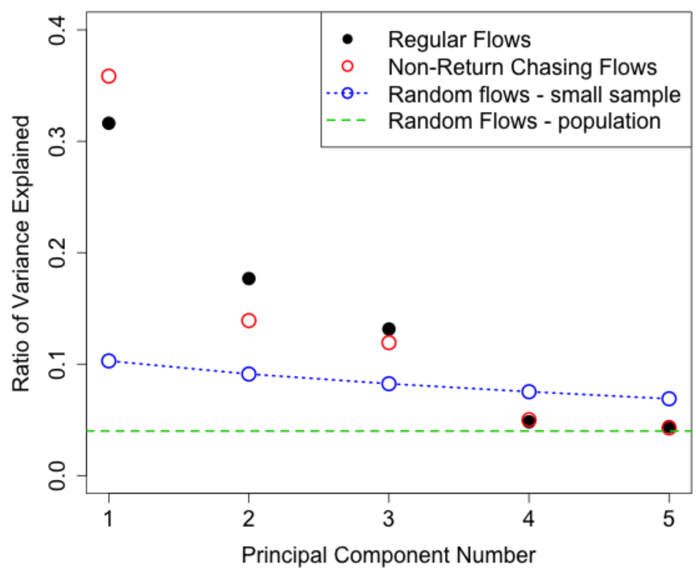

Figure 11. Ratio of variance explained by first principal components of 5 by 5 portfolio flows. Dashed line is what the in-population ratio should be ($1/25 = 4\%$) if flows are random.

	Raw flows						**Non-return chasing flows**				
PC 1											
	Growth	2	3	4	Value		Growth	2	3	4	Value
Big	-0.20	-0.23	-0.22	-0.19	-0.17	Big	-0.21	-0.21	-0.21	-0.18	-0.14
2	-0.20	-0.20	-0.21	-0.23	-0.21	2	-0.19	-0.23	-0.23	-0.21	-0.17
3	-0.23	-0.23	-0.22	-0.26	-0.23	3	-0.23	-0.22	-0.21	-0.25	-0.18
4	-0.19	-0.26	-0.20	-0.17	-0.08	4	-0.20	-0.27	-0.22	-0.20	-0.06
Small	-0.13	-0.22	-0.17	-0.08	-0.14	Small	-0.17	-0.24	-0.20	-0.13	-0.13
PC 2											
	Growth	2	3	4	Value		Growth	2	3	4	Value
Big	-0.32	-0.28	-0.28	-0.12	0.04	Big	-0.32	-0.32	-0.30	-0.17	-0.03
2	-0.22	-0.21	0.00	0.05	0.17	2	-0.20	-0.20	-0.01	0.04	0.16
3	-0.23	-0.03	0.11	0.11	0.20	3	-0.21	0.00	0.12	0.11	0.18
4	-0.11	-0.03	0.16	0.22	0.30	4	-0.03	-0.04	0.14	0.23	0.30
Small	-0.07	0.16	0.28	0.35	0.28	Small	-0.02	0.17	0.30	0.34	0.25
PC 3											
	Growth	2	3	4	Value		Growth	2	3	4	Value
Big	-0.07	-0.13	-0.16	-0.31	-0.33	Big	-0.07	-0.10	-0.12	-0.29	-0.34
2	0.11	-0.05	-0.10	-0.13	-0.29	2	0.10	-0.04	-0.05	-0.17	-0.34
3	0.13	0.22	0.09	-0.04	-0.24	3	0.12	0.21	0.11	-0.08	-0.33
4	0.31	0.11	0.16	0.18	-0.13	4	0.30	0.11	0.12	0.15	-0.18
Small	0.37	0.25	0.24	0.05	-0.23	Small	0.30	0.19	0.18	0.03	-0.30

Table 6. Loadings of first 3 principal components using raw flows (left) or flows after taking out first 4 lags of past fund returns (right).

C.2 Can past factor-level cash flows predict fund flows?

This follows section 3.4 in the main paper.

	Dependent Variable = fund flow$_t$							
	Cash Flow = ROE				Cash Flow = ROA			
	SMB	HML	SMB	HML	SMB	HML	SMB	HML
(Intercept)	0.07*	0.05	0.08*	0.04	0.02	0.02	0.03	0.04
	(0.03)	(0.05)	(0.03)	(0.06)	(0.03)	(0.05)	(0.04)	(0.05)
fund flow$_{t-1}$	0.51***	0.48***	0.65***	0.77***	0.64***	0.51***	0.76***	0.79***
	(0.12)	(0.13)	(0.16)	(0.16)	(0.11)	(0.13)	(0.16)	(0.15)
fund flow$_{t-2}$			−0.17	−0.49*			−0.16	−0.54**
			(0.19)	(0.19)			(0.20)	(0.18)
fund flow$_{t-3}$			−0.17	0.22			−0.06	0.21
			(0.16)	(0.16)			(0.16)	(0.17)
cash flow$_{t-1}$	0.65*	0.48	−0.00	0.49	0.22	0.26	0.88	−0.77
	(0.31)	(0.46)	(0.58)	(0.60)	(1.57)	(1.10)	(2.12)	(1.79)
cash flow$_{t-2}$			0.46	−0.41			−0.24	1.12
			(0.68)	(0.69)			(2.48)	(2.02)
cash flow$_{t-3}$			0.30	0.30			−0.01	0.35
			(0.54)	(0.58)			(2.06)	(1.66)
Adj. R^2	0.43	0.25	0.45	0.33	0.38	0.24	0.38	0.32
N	49	49	47	47	49	49	47	47

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 7. Forecasting regression of annual SMB and HML factor-level fund flows.

References

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer. 2015. X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics* 115:1–24.

Barberis, N., and A. Shleifer. 2003. Style investing. *Journal of financial Economics* 68:161–199.

Blanchard, O. J., and D. Quah. 1989. The Dynamic Effects of Aggregate Demand and Supply Disturbances. *The American Economic Review* 79:655–673.

Bogousslavsky, V. 2016. Infrequent rebalancing, return autocorrelation, and seasonality. *The Journal of Finance* 71:2967–3006.

Campbell, J. Y., and A. S. Kyle. 1993. Smart money, noise trading and stock price behaviour. *The Review of Economic Studies* 60:1–34.

Cella, C., A. Ellul, and M. Giannetti. 2013. Investors' horizons and the amplification of market shocks. *Review of Financial Studies* p. hht023.

Chang, Y.-C., H. Hong, and I. Liskovich. 2014. Regression discontinuity and the price effects of stock market indexing. *Review of Financial Studies* p. hhu041.

Chetty, R. 2006. A new method of estimating risk aversion. *The American Economic Review* 96:1821–1834.

Cohen, R. B., C. Polk, and T. Vuolteenaho. 2003. The value spread. *The Journal of Finance* 58:609–642.

Coval, J., and E. Stafford. 2007. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics* 86:479–512.

Daniel, K., and S. Titman. 1997. Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance* 52:1–33.

Duffie, D. 2010. Presidential Address: Asset Price Dynamics with Slow-Moving Capital. *The Journal of finance* 65:1237–1267.

Edmans, A., I. Goldstein, and W. Jiang. 2012. The real effects of financial markets: The impact of prices on takeovers. *The Journal of Finance* 67:933–971.

Fama, E. F., and K. R. French. 1995. Size and book-to-market factors in earnings and returns. *The Journal of Finance* 50:131–155.

Fama, E. F., and K. R. French. 1996. Multifactor explanations of asset pricing anomalies. *The journal of finance* 51:55–84.

Frazzini, A., and O. A. Lamont. 2008. Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics* 88:299–322.

Greenwood, R. 2005. Short-and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics* 75:607–649.

Greenwood, R., and S. G. Hanson. 2012. Share issuance and factor timing. *The Journal of Finance* 67:761–798.

Greenwood, R., S. G. Hanson, and G. Y. Liao. 2016. Asset price dynamics in partially segmented markets. Tech. rep.

Greenwood, R., and D. Vayanos. 2014. Bond supply and excess bond returns. *The Review of Financial Studies* 27:663–713.

Hodrick, R. J. 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial studies* 5:357–386.

Kozak, S., S. Nagel, and S. Santosh. 2017. Interpreting factor models .

Kyle, A. S. 1985. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society* pp. 1315–1335.

Lou, D. 2012. A flow-based explanation for return predictability. *Review of financial studies* 25:3457–3489.

Morningstar. 2016. Morningstar Style Box Methodology. `http://hk.morningstar.com/ODS_Images/Morningstar_StyleBox_Methodology.pdf`. Accessed: 2016-03-10.

Petajisto, A. 2013. Active share and mutual fund performance. *Financial Analysts Journal* 69:73–93.

Shleifer, A. 1986. Do demand curves for stocks slope down? *The Journal of Finance* 41:579–590.

Wermers, R. 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *The Journal of Finance* 55:1655–1695.