Online Appendix

Fund Flows, Liquidity, and Asset Prices

Minsoo Kim

A. Proofs

A.1. Proof of Proposition 1

I start from household i's portfolio choice problem

$$\max_{\mathbf{x}_{i,t}} \quad \mathbb{E}_t\left[\mathbf{x}_{i,t}^T\left(\mathbf{r}_{t+1}^c - r_f\mathbf{1}\right)\right] - \frac{b_I}{2}Var_t\left[\mathbf{x}_{i,t}^T\left(\mathbf{r}_{t+1}^c - r_f\mathbf{1}\right)\right]. \tag{OA.1}$$

I take the first order condition and rearrange the terms to get

$$\mathbf{x}_{i,t}^* = \frac{1}{b_I}\Sigma_t^{-1}\left(\mathbb{E}_t\left[\mathbf{r}_{t+1}^c\right] - r_f\mathbf{1}\right). \tag{OA.2}$$

For mutual fund manager j, the portfolio choice problem is

$$\max_{\mathbf{x}_{j,t}} \mathbb{E}_t\left[\mathbf{x}_{j,t}^T\mathbf{r}_{t+1}^c + f_{j,t+1}\right] - \frac{b_J}{2}Var_t\left[\mathbf{x}_{j,t}^T\mathbf{r}_{t+1}^c + f_{j,t+1}\right] - \lambda\left(\mathbf{x}_{j,t}^T\mathbf{1} - 1\right). \tag{OA.3}$$

Substituting the fund flows in equation (A.7) into equation (OA.3), manager j's maximization problem becomes

$$\begin{aligned}
&\max_{\mathbf{x}_{j,t}} \mathbb{E}_t\left[\alpha_0 + \left\{(1+\alpha_1)\mathbf{x}_{j,t} - \alpha_1 x_{M,t}\right\}^T\mathbf{r}_{t+1}^c + \eta_{j,t+1}\right] \\
-\ &\frac{b_J}{2}Var_t\left[\left\{(1+\alpha_1)\mathbf{x}_{j,t} - \alpha_1 x_{M,t}\right\}^T\mathbf{r}_{t+1}^c + \eta_{j,t+1}\right] - \lambda\left(\mathbf{x}_{j,t}\mathbf{1} - 1\right).
\end{aligned} \tag{OA.4}$$

For notational simplicity, I denote

$$\mathbb{E}_t \equiv \mathbb{E}_t\left[\mathbf{r}_{t+1}^c\right]_{K\times 1,} \tag{OA.5}$$

$$\Sigma_t \equiv Var_t\left[\mathbf{r}_{t+1}^c\right]_{K\times K,} \tag{OA.6}$$

$$C_{j,t} \equiv Cov_t\left[\mathbf{r}_{t+1}^c, \eta_{j,t+1}\right]_{K\times 1.} \tag{OA.7}$$

I rewrite manager j's problem in a shorter expression after expanding the variance term

$$\begin{aligned}
&\max_{\mathbf{x}_{j,t}} \left\{(1+\alpha_1)\mathbf{x}_{j,t} - \alpha_1\mathbf{x}_{M,t}\right\}^T\mathbb{E}_t - \frac{b_J}{2}\left\{(1+\alpha_1)\mathbf{x}_{j,t} - \alpha_1\mathbf{x}_{M,t}\right\}^T\Sigma_t\left\{(1+\alpha_1)\mathbf{x}_{j,t} - \alpha_1\mathbf{x}_{M,t}\right\} \\
&- b_J\left\{(1+\alpha_1)\mathbf{x}_{j,t} - \alpha_1\mathbf{x}_{M,t}\right\}^T C_{j,t} - \frac{b_J}{2}Var_t[\eta_{j,t+1}] - \lambda\left(\mathbf{x}_{j,t}^T\mathbf{1} - 1\right).
\end{aligned} \tag{OA.8}$$

The first order condition with respect to $\mathbf{x}_{j,t}$ reads

$$(1+\alpha_1)\mathbb{E}_t - b_J(1+\alpha_1)^2\Sigma_t\mathbf{x}_{j,t} + b_J\Sigma_t(1+\alpha_1)\alpha_1\mathbf{x}_{M,t} - b_J(1+\alpha_1)C_{j,t} - \lambda\mathbf{1} = 0. \tag{OA.9}$$

After rearranging the terms, the optimal portfolio weight of fund manager j becomes

$$\mathbf{x}^*_{j,t} = \frac{\alpha_1}{1+\alpha_1}\mathbf{x}_{M,t} + \frac{1}{b_J(1+\alpha_1)}\Sigma_t^{-1}\left(\mathbb{E}_t - \frac{1}{1+\alpha_1}\lambda\mathbf{1}\right) - \frac{1}{1+\alpha_1}\Sigma_t^{-1}C_{j,t}. \tag{OA.10}$$

I replace $\mathbf{x}^*_{i,t}$ and $\mathbf{x}^*_{j,t}$ in the market clearing condition

$$\sum_{i=1}^{I} A_{i,t}\mathbf{x}^*_{i,t} + \sum_{j=1}^{J} A_{j,t}\mathbf{x}^*_{j,t} = A_{M,t}\mathbf{x}_{M,t}, \tag{OA.11}$$

to get

$$\sum_{j=1}^{J} A_{j,t}\left\{\frac{\alpha_1}{1+\alpha_1}\mathbf{x}_{M,t} + \frac{1}{b_J(1+\alpha_1)}\Sigma_t^{-1}\left(\mathbb{E}_t - \frac{1}{1+\alpha_1}\lambda\mathbf{1}\right) - \frac{1}{1+\alpha_1}\Sigma_t^{-1}C_{j,t}\right\}$$
$$+ \sum_{i=1}^{I} A_{i,t}\left\{\frac{1}{b_I}\Sigma_t^{-1}\left(\mathbb{E}_t - r_f\mathbf{1}\right)\right\} = A_{M,t}\mathbf{x}_{M,t}. \tag{OA.12}$$

I pre-multiply by Σ_t on both sides of the equation

$$\sum_{j=1}^{J}\frac{\alpha_1 A_{j,t}}{1+\alpha_1}\Sigma_t\mathbf{x}_{M,t} + \sum_{j=1}^{J}\frac{A_{j,t}}{b_J(1+\alpha_1)}\left(\mathbb{E}_t - \frac{1}{1+\alpha_1}\lambda\mathbf{1}\right) - \sum_{j=1}^{J}\frac{A_{j,t}}{1+\alpha_1}C_{j,t} + \sum_{i=1}^{I}\frac{A_{i,t}}{b_I}\left(\mathbb{E}_t - r_f\mathbf{1}\right)$$
$$= A_{M,t}\Sigma_t\mathbf{x}_{M,t}, \tag{OA.13}$$

and rearrange using $\sum_{i=1}^{I} A_{i,t} \equiv A^{total}_{I,t}$, $\sum_{j=1}^{J} A_{j,t} \equiv A^{total}_{J,t}$, and $A_{M,t} \equiv A^{total}_{I,t} + A^{total}_{J,t}$

$$\left(\frac{1}{b_J(1+\alpha_1)}A^{total}_{J,t} + \frac{1}{b_I}A^{total}_{I,t}\right)\mathbb{E}_t \tag{OA.14}$$
$$= \frac{1}{b_J(1+\alpha_1)^2}A^{total}_{J,t}\lambda 1 + \frac{1}{b_I}A^{total}_{I,t}r_f\mathbf{1} + A_{M,t}\Sigma_t\mathbf{x}_{M,t} - \frac{\alpha_1}{1+\alpha_1}A^{total}_{J,t}\Sigma_t\mathbf{x}_{M,t} + \frac{1}{1+\alpha_1}\sum_{j=1}^{J}A_{j,t}C_{j,t}.$$

I divide both sides by $\left(\frac{1}{b_J(1+\alpha_1)}A^{total}_{J,t} + \frac{1}{b_I}A^{total}_{I,t}\right)$ and restore the original expressions for $\mathbb{E}_t$, Σ_t, and $C_{j,t}$ to get

$$\mathbb{E}_t\left[\mathbf{r}^c_{t+1}\right] = s_0\mathbf{1} + s_1 Var_t\left[\mathbf{r}^c_{t+1}\right]\mathbf{x}_{M,t} + s_2 Cov_t\left[\mathbf{r}^c_{t+1}, \bar{\eta}_{t+1}\right], \tag{OA.15}$$

and I rewrite for each security $k \in \{1, 2, 3, ..., K\}$

$$\mathbb{E}_t\left[r^c_{k,t+1}\right] = s_0 + s_1 Cov_t(r^c_{k,t+1}, r^c_{M,t+1}) + s_2 Cov_t(r^c_{k,t+1}, \bar{\eta}_{t+1}), \tag{OA.16}$$

where

$$\bar{\eta}_{t+1} = \sum_{j=1}^{J} \frac{A_{j,t}}{A_{J,t}^{total}} \eta_{j,t+1}, \tag{OA.17}$$

$$s_0 = r_f \frac{\frac{1}{b_I} A_{I,t}^{total}}{\frac{1}{b_J(1+\alpha_1)} A_{J,t}^{total} + \frac{1}{b_I} A_{I,t}^{total}} + \lambda \frac{\frac{1}{b_J(1+\alpha_1)^2} A_{J,t}^{total}}{\frac{1}{b_J(1+\alpha_1)} A_{J,t}^{total} + \frac{1}{b_I} A_{I,t}^{total}}, \tag{OA.18}$$

$$s_1 = \frac{A_{M,t} - \frac{\alpha_1}{1+\alpha_1} A_{J,t}^{total}}{\frac{1}{b_J(1+\alpha_1)} A_{J,t}^{total} + \frac{1}{b_I} A_{I,t}^{total}}, \qquad s_2 = \frac{\frac{1}{1+\alpha_1} A_{J,t}^{total}}{\frac{1}{b_J(1+\alpha_1)} A_{J,t}^{total} + \frac{1}{b_I} A_{I,t}^{total}}. \tag{OA.19}$$

A.2. Proof of Proposition 2

I start from household i's portfolio choice problem

$$\max_{\mathbf{x}_{i,t}} \quad \mathbb{E}_t \left[\mathbf{x}_{i,t}^T \left(\mathbf{r}_{t+1}^c - r_f \mathbf{1} \right) \right] - \frac{b_I}{2} Var_t \left[\mathbf{x}_{i,t}^T \left(\mathbf{r}_{t+1}^c - r_f \mathbf{1} \right) \right], \tag{OA.20}$$

and take the first order condition and rearrange the terms to get

$$\mathbf{x}_{i,t}^* = \frac{1}{b_I} \Sigma_t^{-1} \left(\mathbb{E}_t \left[\mathbf{r}_{t+1}^c \right] - r_f \mathbf{1} \right). \tag{OA.21}$$

I replace $\mathbf{x}_{i,t}^*$ in the market clearing condition

$$\sum_{i=1}^{I} A_{i,t} \mathbf{x}_{i,t}^* = A_{M,t} \mathbf{x}_{M,t}, \tag{OA.22}$$

to get

$$\sum_{i=1}^{I} A_{i,t} \left\{ \frac{1}{b_I} \Sigma_t^{-1} \left(\mathbb{E}_t \left[\mathbf{r}_{t+1}^c \right] - r_f \mathbf{1} \right) \right\} = A_{M,t} \mathbf{x}_{M,t}. \tag{OA.23}$$

I pre-multiply by Σ_t on both sides of the equation

$$\sum_{i=1}^{I} \frac{A_{i,t}}{b_I} \left(\mathbb{E}_t \left[\mathbf{r}_{t+1}^c \right] - r_f \mathbf{1} \right) = A_{M,t} \Sigma_t \mathbf{x}_{M,t}, \tag{OA.24}$$

and rearrange using $\sum_{i=1}^{I} A_{i,t} \equiv A_{I,t}^{total}$ and $A_{M,t} \equiv A_{I,t}^{total}$

$$\frac{1}{b_I} A_{I,t}^{total} \mathbb{E}_t \left[\mathbf{r}_{t+1}^c \right] = \frac{1}{b_I} A_{I,t}^{total} r_f \mathbf{1} + A_{I,t}^{total} \Sigma_t \mathbf{x}_{M,t}. \tag{OA.25}$$

I divide both sides by $\frac{1}{b_I} A_{I,t}^{total}$

$$\mathbb{E}_t \left[\mathbf{r}_{t+1}^c \right] = r_f \mathbf{1} + b_I Var_t \left[\mathbf{r}_{t+1}^c \right] \mathbf{x}_{M,t}, \tag{OA.26}$$

and I rewrite for each security $k \in \{1, 2, 3, ..., K\}$

$$\mathbb{E}_t \left[r_{k,t+1}^c \right] = r_f + b_I Cov_t (r_{k,t+1}^c, r_{M,t+1}^c). \tag{OA.27}$$

Rearranging the terms yields the liquidity-adjusted CAPM of Acharya and Pedersen (2005)

$$\mathbb{E}_t\left[r^c_{k,t+1}\right] - r_f = \frac{Cov_t(r^c_{k,t+1}, r^c_{M,t+1})}{Var_t(r^c_{k,t+1})}\mathbb{E}_t\left[r^c_{k,t+1} - r_f\right]. \tag{OA.28}$$

Rewriting equation (OA.20) in terms of gross returns, $r_{k,t+1}$, and repeating the same derivation yield the CAPM as in Sharpe (1964) and Lintner (1965).

A.3. Proof of Proposition 3

An unconditional version of the two-factor model in Proposition 1 can be derived under the assumption that the dividends and the liquidity costs in equation (A.1) are independent over time, that is $\rho^D = \rho^C = 0$. Empirically, however, liquidity costs are persistent over time, thus I instead assume that the conditional covariances in the asset pricing model are constant over time. An alternative method that gives the same unconditional model is to assume the risk prices are constant over time and use the fact that for any random variable X and Y, $\mathbb{E}\left[\text{Cov}_t(X,Y)\right] = \text{Cov}(X - \mathbb{E}_t(X), Y - \mathbb{E}_t(Y))$. The unconditional version of the equilibrium model for each security k is then written as

$$\mathbb{E}\left[r^c_{k,t}\right] = s_0 + s_1\text{Cov}\left(r_{k,t} - \varepsilon^c_{k,t}, \varepsilon^r_{M,t} - \varepsilon^c_{M,t}\right) + s_2\text{Cov}\left(r_{k,t} - \varepsilon^c_{k,t}, \varepsilon^{\bar{\eta}}_t\right), \tag{OA.29}$$

where $\varepsilon^c_{k,t} \equiv c_{k,t} - E_{t-1}[c_{k,t}]$, $\varepsilon^{\bar{\eta}}_t \equiv \bar{\eta}_t - E_{t-1}[\bar{\eta}_t]$, $\varepsilon^r_{M,t} \equiv r_{M,t} - E_{t-1}[r_{M,t}]$ and $\varepsilon^c_{M,t} \equiv c_{M,t} - E_{t-1}[c_{M,t}]$.

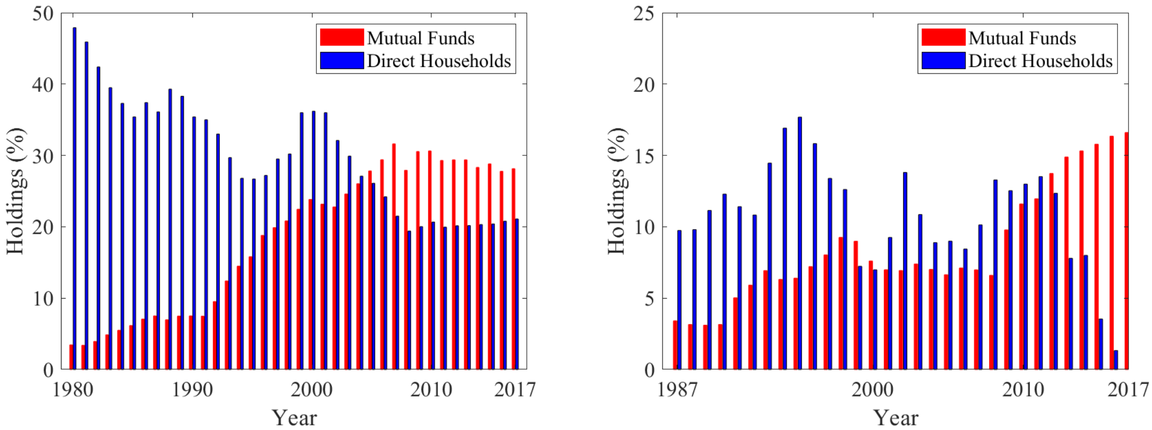
B. Supplementary Figures and Tables

Figure OA.1: Shares of Holdings by Mutual Funds and Direct Households

In Panel (a), I plot the shares in the U.S. common equity market held by direct households and open-end mutual funds from 1980 to 2017. The data is from the Federal Reserve Board Flow of Funds Accounts L.223, which reports the dollar values of the corporate equity held by various types of investors, including households and nonprofits, mutual funds, banks, and insurance companies.

In the Flow of Funds, the household and nonprofit holdings are computed as a residual after subtracting the holdings of other sectors, and include not only the common equity held by the nonprofit sector, but also the value of preferred stocks and closely held corporations. Similarly to Stambaugh (2014), I obtain the shares of the direct household holdings between 1980 and 2007 from French (2008) and estimate the values of the direct household holdings from 2008 onward by using the ratio of the direct household holdings over the total household and nonprofit holdings in 2007 from French (2008). I extrapolate the ratio to extend the estimates of the direct household holdings to 2017.

In Panel (b), I plot the shares of the U.S. corporate and foreign bonds held by direct households and open-end mutual funds from 1987 to 2017. The dollar values of the mutual fund holdings and direct household holdings are from the Federal Reserve Board Flow of Funds Accounts L.213 and B.101.h, respectively. The plot starts from 1987 because direct household holdings (B.101.h) are only available from 1987. The Flow of Funds Account does not provide separate dollar values of U.S. corporate bonds excluding foreign bonds. Direct household holdings (B.101.h) do not include non-profit holdings (B.101.n). Note that the direct household holdings provide an upper bound of true direct household shares because it is computed as a residual.

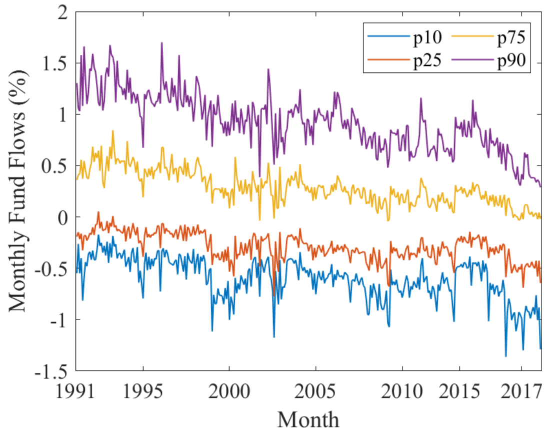


(a) Common Equity Market

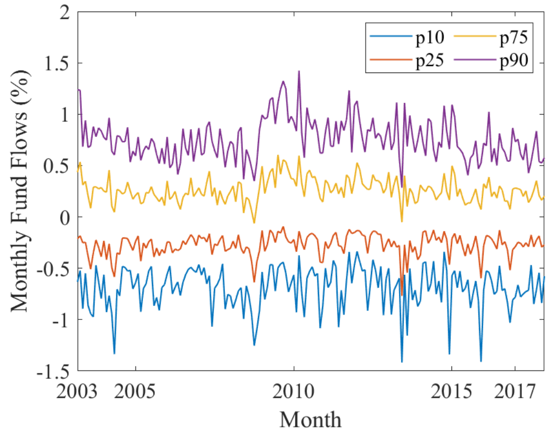
(b) Corporate Bond Market

Figure OA.2: Common Movements of Individual Fund Flow Shocks

Figure (a) plots the cross-sectional interquartile range and interdecile range of the individual fund flow shocks in each month for equity funds between February 1991 and December 2017, with corporate bond funds between February 2003 and December 2017 in figure (b). I run a panel regression of fund flows on the lagged fund flows and the fund's past performance over the market benchmark with time-fixed effects using all qualified mutual funds. I then compute the individual shocks to fund flows as the sum of the time-fixed effects and the idiosyncratic flow shocks as in equation (5). For display purposes, I divide the interquartile range and interdecile range by its time-series standard deviation over the sample period.



(a) Equity Funds



(b) Bond Funds

Figure OA.3: Cross-Sectional Fit of 150 Test Portfolios

The plot shows the cross-sectional fit of the asset pricing model using 100 stock portfolios and 50 corporate bond portfolios. In each portfolio $p \in \{1, 2, ..., 150\}$, I take the value-weighted average of excess returns over the 30 day T-bill rate across assets and then take the time-series average of the portfolio returns. I compute the fitted monthly excess returns using the risk prices estimated from 100 stock portfolios and 50 bond portfolios, reported under the column All in Table 2. For the 25 liquidity and fund flow beta portfolios (LiqFlow), the label ij refers to the portfolio of stocks/bonds with liquidity costs in quintile i and return beta on flows in quintile j for $i, j \in \{1, 2, ..., 5\}$. For the 25 book-to-market by size (B/M) portfolios, the 25 profitability by size (Profit) portfolios, and the 25 investment by size (Invest) portfolios, the label ij indicates the portfolio of stocks with a size in quintile i and book-to-market, profitability, and investment, respectively, in quintile j for $i, j \in \{1, 2, ..., 5\}$. For the 25 credit rating and maturity portfolios (CreditMaturity), the label ij indicates the portfolio of bonds with a credit rating in quintile i and maturity in quintile j for $i, j \in \{1, 2, ..., 5\}$.

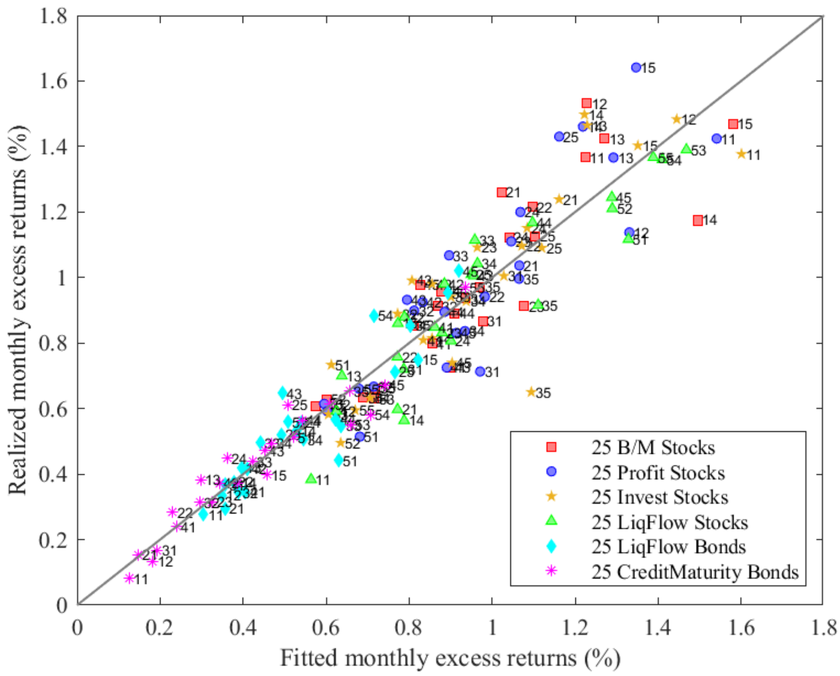
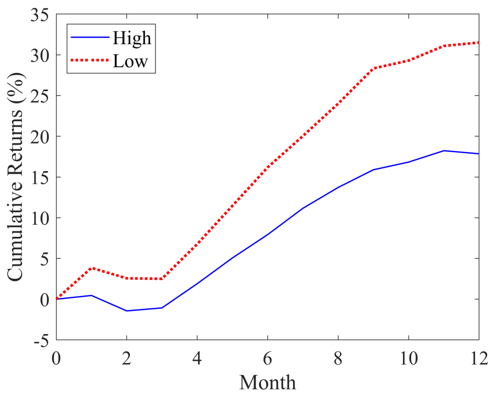


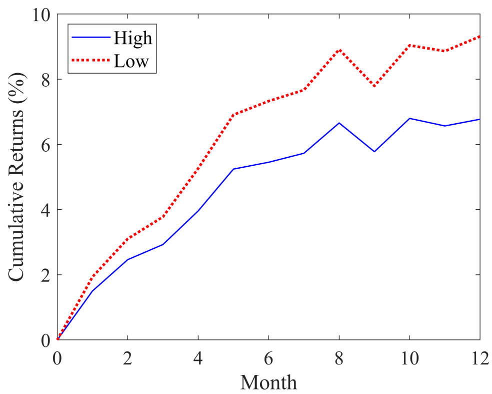
Figure OA.4: Shocks to Fund Flow Betas and Cumulative Returns Over Time

In Panel A, I plot the cumulative bond returns of portfolios sorted on the liquidity beta on flows following the 2008 financial crisis in figure (a) and following the 2013 taper tantrum shocks in figure (b), with Low referring to the bottom-quartile portfolio with the most negative liquidity beta on flows. In Panel B, I plot the cumulative bond returns of portfolios sorted on the return beta on flows following the 2008 financial crisis in figure (c) and following the 2013 taper tantrum shocks in figure (d), with High referring to the top-quartile portfolio with a high return beta on flows.

Panel A: Liquidity Beta on Flows

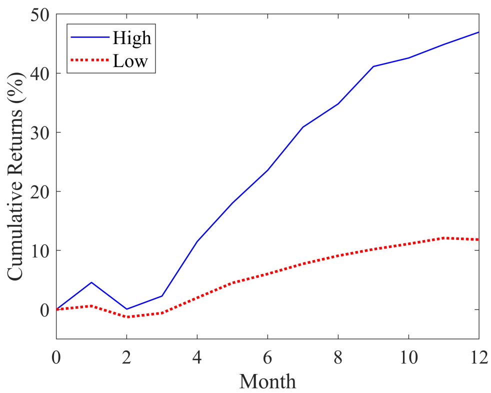


(a) 2008 Financial Crisis

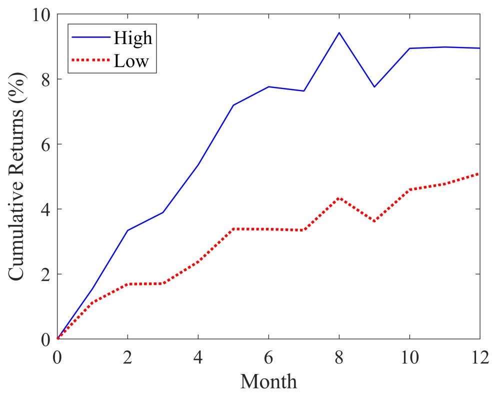


(b) 2013 Taper Tantrum

Panel B: Return Beta on Flows



(c) 2008 Financial Crisis



(d) 2013 Taper Tantrum

Table OA.1: Cross-Sectional Asset Pricing Tests Using Test Portfolios

This table reports the prices of risk for the fund flow beta and liquidity-adjusted market beta in the cross-sectional tests of the asset pricing model in Proposition 3. I impose an economic restriction that investors pay only the monthly portion of the liquidity costs. Specifically, I calibrate the unknown holding period using the sample average monthly turnovers of 0.19 from all stocks in the test portfolios. Fund flow beta, β^{FLOW}, equals the return beta on flows minus the liquidity beta on flows, assuming a common risk price. The return (liquidity) beta on flows captures the co-movement of returns (liquidity costs) with the aggregate fund flow shocks. The liquidity-adjusted market beta, β^{LMKT}, is the co-movement of net returns, i.e., returns minus liquidity costs, with the net market returns. $\mathbb{E}[c]$ is the expected level of liquidity costs. LiqFlow refers to 25 liquidity and fund flow beta, B/M denotes 25 book-to-market by size, Profit denotes 25 profitability by size, Invest denotes 25 investment by size portfolios, and All refers to the union of 100 test portfolios. The test portfolios are formed using NYSE stocks excluding financial firms. The monthly sample period is from January 1996 to December 2017. Mean absolute pricing errors (MAPE) is the cross-sectional average of the absolute difference between actual monthly excess returns and model-predicted monthly excess returns. GMM t-statistics in parentheses are computed using the Newey and West (1987) method with two lags to take into account the pre-estimation of the time-series betas and serial correlations. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	LiqFlow	B/M	Profit	Invest	All
Flow β^{FLOW}	0.20***	0.10**	0.25***	0.15**	0.17***
	(4.04)	(1.98)	(2.64)	(2.44)	(4.69)
Market β^{LMKT}	-0.05	0.64**	-0.15	0.09	0.17
	(-0.18)	(2.36)	(-0.24)	(0.15)	(0.78)
Intercept α	0.03	-0.14	-0.03	0.12	-0.01
	(0.18)	(-0.81)	(-0.11)	(0.52)	(-0.14)
Liquidity Costs $\mathbb{E}[c]$	0.19	0.19	0.19	0.19	0.19
	(–)	(–)	(–)	(–)	(–)
R^2	0.72	0.62	0.64	0.50	0.60
Adj. R^2	0.69	0.59	0.61	0.46	0.59
MAPE (%)	0.17	0.14	0.16	0.18	0.15
Assets	25	25	25	25	100
Months	264	264	264	264	264

Table OA.2: Characteristics of 100 Stock Portfolios

This table presents the characteristics of the 100 stock portfolios in Table 2. A full description of the test portfolio construction is in Online Appendix D. In sub-panel (a), the fund flow beta β^{FLOW} is the return beta on flows minus the liquidity beta on flows. The return (liquidity) beta on flows captures the co-movement of returns (liquidity costs) with the aggregate fund flow shocks. In sub-panel (b), the portfolio-level returns are the value-weighted averages of the returns in excess of the one-month T-bill rate, using the market capitalization as the weight.

Panel A: 25 Liquidity and Flow Beta Portfolios

(a) β^{FLOW}

	Flow beta				
	Low	2	3	4	High
Liq	2.38	2.75	2.86	3.60	3.60
2	3.60	3.58	4.19	4.29	4.55
3	3.65	3.57	4.67	4.55	5.47
4	3.88	4.00	4.37	5.23	6.23
Illiq	4.69	3.77	4.71	4.39	4.83

(b) Excess returns (%)

	Flow beta				
	Low	2	3	4	High
Liq	0.38	0.58	0.72	0.59	0.83
2	0.59	0.75	0.85	0.79	1.02
3	0.73	0.87	1.11	1.04	0.93
4	0.85	0.99	0.99	1.18	1.24
Illiq	1.12	1.21	1.38	1.38	1.36

Panel B: 25 Book-to-Market by Size

(a) β^{FLOW}

	Book-to-market				
	Low	2	3	4	High
Small	4.70	4.68	4.87	5.73	5.63
2	4.74	5.15	5.04	4.85	5.19
3	4.69	4.05	4.45	4.11	4.67
4	4.04	3.80	4.30	4.37	3.90
Big	2.45	2.60	3.27	3.12	3.30

(b) Excess returns (%)

	Book-to-market				
	Low	2	3	4	High
Small	1.37	1.53	1.42	1.18	1.47
2	1.26	1.22	0.91	1.12	1.13
3	0.87	0.91	0.94	0.96	0.97
4	0.80	0.85	0.73	0.89	0.98
Big	0.61	0.63	0.63	0.63	0.66

Panel C: 25 Profitability by Size

(a) β^{FLOW}

	Profitability				
	Low	2	3	4	High
Small	5.75	4.72	5.01	4.62	5.21
2	4.91	4.49	4.89	4.99	5.54
3	4.63	3.74	4.23	4.44	5.18
4	4.24	3.93	3.70	4.22	4.38
Big	3.05	3.26	2.57	2.62	3.06

(b) Excess returns (%)

	Profitability				
	Low	2	3	4	High
Small	1.42	1.14	1.37	1.46	1.64
2	1.04	0.94	1.11	1.20	1.43
3	0.71	0.90	1.07	0.84	1.00
4	0.73	0.92	0.93	0.90	0.83
Big	0.51	0.67	0.61	0.59	0.66

Panel D: 25 Investment by Size

(a) β^{FLOW}

	Investment				
	Low	2	3	4	High
Small	5.97	5.32	4.34	4.54	5.33
2	5.49	4.96	4.37	5.08	5.28
3	4.97	3.51	4.25	4.45	5.35
4	3.92	4.05	3.77	4.04	4.31
Big	2.68	2.81	2.65	3.20	2.98

(b) Excess returns (%)

	Investment				
	Low	2	3	4	High
Small	1.38	1.48	1.46	1.50	1.40
2	1.24	1.10	1.09	1.15	1.09
3	1.00	0.89	0.94	0.93	0.65
4	0.81	0.98	0.99	0.82	0.74
Big	0.73	0.50	0.58	0.63	0.60

Table OA.3: Characteristics of 50 Bond Portfolios

This table presents the characteristics of the 50 corporate bond portfolios in Table 2. A full description of the test portfolio construction is in Online Appendix D. In sub-panel (a), the fund flow beta β^{FLOW} is the return beta on flows minus the liquidity beta on flows. The return (liquidity) beta on flows captures the co-movement of returns (liquidity costs) with the aggregate fund flow shocks. In sub-panel (b), the portfolio-level returns are the value-weighted averages of the returns in excess of the one-month T-bill rate, using the market capitalization as the weight.

Panel A: 25 Liquidity and Flow Beta Portfolios

(a) β^{FLOW}

	Flow beta				
	Low	2	3	4	High
Liq	0.79	0.91	1.13	1.57	2.92
2	1.00	1.07	1.58	2.01	3.12
3	0.87	1.10	1.29	1.97	3.07
4	1.03	1.07	1.50	2.09	3.51
Illiq	1.47	1.08	1.57	2.09	2.74

(b) Excess returns (%)

	Flow beta				
	Low	2	3	4	High
Liq	0.28	0.34	0.42	0.53	0.75
2	0.29	0.38	0.52	0.56	0.71
3	0.37	0.34	0.50	0.51	0.85
4	0.35	0.42	0.65	0.57	1.02
Illiq	0.44	0.56	0.55	0.88	0.95

Panel B: 25 Credit Rating and Maturity Portfolios

(a) β^{FLOW}

	Maturity				
	Short	2	3	4	Long
AAA	0.01	0.12	0.49	0.68	0.47
AA	0.14	0.46	0.81	0.90	1.31
A	0.34	0.75	1.18	1.30	1.73
BBB	0.50	1.00	1.39	1.71	2.37
HY	1.68	2.13	2.57	2.78	3.53

(b) Excess returns (%)

	Maturity				
	Short	2	3	4	Long
AAA	0.08	0.13	0.38	0.37	0.40
AA	0.15	0.28	0.31	0.45	0.61
A	0.17	0.31	0.44	0.49	0.65
BBB	0.24	0.37	0.47	0.56	0.67
HY	0.51	0.61	0.55	0.58	0.97

Table OA.4: Time-Series Analysis of Factors Underlying the Liquidity Beta on Flows

This table presents panel regressions with bond fixed effects to examine time-series relationships underlying the liquidity beta on flows. Panel A estimates the effect of aggregate fund flow shocks on two measures of mutual fund trading activity – Mutual Funds Trading Share and an indicator for severe selling – including bond fixed effects and controls. Panel B then regresses bond liquidity costs on the two measures of mutual fund trading activity from Panel A, including bond fixed effects and controls. The sample period is from Q1.2008 to Q4.2017. Mutual Funds Trading Share is the aggregate trading volume by all active corporate bond mutual funds as a fraction of the total trading volume of the bond. The time-series indicator, 1{Severe Selling}, equals one if a bond-quarter is in the top decile, within the bond's time series, of mutual funds' aggregate selling as a fraction of the bond's total trading volume, and zero otherwise. To avoid look-ahead bias, I compute the liquidity beta on flows using only the data available at the end of each quarter in the preceding 60-month rolling windows. Controls in Columns (2) and (4) include Maturity and Size. Maturity is the number of years to maturity, and Size is the log of the market capitalization, as of the preceding month. Standard errors are two-way clustered at bond and year-quarter level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Links Between Aggregate Fund Flows and Mutual Fund Trading

	Mutual Funds Trading Share		1{Severe Selling}	
	(1)	(2)	(3)	(4)
Aggregate Fund Flow Shocks	-1.098*	-1.057*	-0.006**	-0.006**
	(-1.99)	(-1.97)	(-2.19)	(-2.21)
Controls	No	Yes	No	Yes
Bond F. E.	Yes	Yes	Yes	Yes
Observations	71,721	71,721	14,307	14,307

Panel B: Links Between Mutual Fund Trading and Liquidity Costs

	Liquidity Costs		Liquidity Costs	
	(1)	(2)	(3)	(4)
Mutual Funds Trading Share	0.034***	0.032***		
	(7.27)	(7.22)		
1{Severe Selling}			0.035***	0.031***
			(3.72)	(3.57)
Controls	No	Yes	No	Yes
Bond F. E.	Yes	Yes	Yes	Yes
Observations	71,607	71,211	14,298	14,282

Table OA.5: Characterization of the Aggregate Shocks to Fund Flows

The table documents the results of linear regressions of the aggregate shocks to fund flows on various aggregate economic variables from February 1991 to December 2017 for equity mutual funds in Panel A and from February 2003 to December 2017 for corporate bond mutual funds in Panel B. VIX is the AR(1) residual of the Chicago Board Options Exchange (CBOE) Volatility Index obtained from the CBOE. Macro Uncertainty is the AR(1) residual of the macroeconomic uncertainty of Jurado, Ludvigson, and Ng (2015). Bearish Sentiment is the AR(1) residual of the market sentiment index, computed as the percentage of survey respondents that say the market will trend down over the next six months, obtained from the American Association of Individual Investors (AAII) webpage. BAB is the betting-against-beta factor (Frazzini and Pedersen, 2014), obtained from the AQR webpage. TED is the spread between the three-month LIBOR and the three-month Treasury bill rates, obtained from the Federal Reserve Economic Data (FRED). t-statistics in parentheses are computed using the Newey and West (1987) method to take into account serial correlations. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Bond Funds

	Dependent Variable: Aggregate Shocks to Fund Flows						
	VIX	Macro Uncertainty	Bearish Sentiment	BAB	TED	Market Returns	Market Illiquidity
	-0.05**	-0.14***	-0.01	0.03	-0.37**	0.23***	-7.77***
	(-2.51)	(-3.38)	(-1.03)	(1.32)	(-2.11)	(5.15)	(-5.69)
R-squared	0.085	0.127	0.006	0.019	0.056	0.240	0.235
Observations	179	179	179	179	179	179	179

Panel B: Equity Funds

	Dependent Variable: Aggregate Shocks to Fund Flows						
	VIX	Macro Uncertainty	Bearish Sentiment	BAB	TED	Market Returns	Market Illiquidity
	-0.04***	-0.04	-0.01***	0.02**	-0.34***	0.03***	-0.10
	(-5.31)	(-1.64)	(-3.18)	(1.94)	(-5.18)	(3.99)	(-0.29)
R-squared	0.091	0.014	0.033	0.016	0.071	0.068	0.000
Observations	323	323	323	323	323	323	323

Table OA.6: Granularity of Bond Fund Flows

This table reports the granularity of the monthly fund flow shocks in active corporate bond funds between January 2003 and December 2017. The fund flow shocks are the residuals of the monthly fund flows after deducting the predicted flows from the past fund performance and the lagged flows as in equation (5). In Panel A, I focus on the months of large aggregate outflow shocks. Monthly Net Flow Shocks is the sum of the signed monthly fund flow shocks across all active corporate bond funds in the year-month: a negative (positive) sign indicates outflow (inflow) shocks. In Panel B, Monthly Absolute Flows Shocks is the sum of the absolute value of the monthly flow shocks across all active corporate bond funds averaged within the year. The Herfindahl–Hirschman Index is the sum of the squared share of the absolute monthly flow shocks across all active corporate bond funds averaged within the year.

Panel A: Share of Flow Shocks by the Largest 20% of Funds During Outflows

Year-Month	Share by the Top 20% Funds (%)	Share by the Rest 80% Funds (%)	Monthly Net Flow Shocks (in $mil.)
2008 October	79	21	-8,279
2011 August	83	17	-4,664
2013 June	85	15	-20,596
2015 December	72	28	-7,801
2016 November	82	18	-9,905

Panel B: Share of Absolute Flow Shocks by the Largest 20% of Funds

Year	Share by the Top 20% Funds (%)	Share by the Rest 80% Funds (%)	Monthly Absolute Flow Shocks (in $mil.)	Herfindahl–Hirschman Index (%)
2003	62	38	3,956	1.92
2004	60	40	3,453	1.84
2005	58	42	3,301	1.75
2006	62	38	3,983	2.19
2007	67	33	5,142	2.80
2008	69	31	5,834	2.12
2009	73	27	8,081	3.61
2010	69	31	8,582	1.96
2011	70	30	9,845	3.16
2012	75	25	11,309	2.74
2013	74	26	12,802	3.25
2014	75	25	13,779	3.15
2015	75	25	13,443	3.41
2016	76	24	13,267	2.98
2017	77	23	12,619	3.11

Table OA.7: Granularity of Fund Flows

This table reports the Herfindahl–Hirschman Index of monthly fund flow shocks in active equity mutual funds between February 1991 and December 2017 and in corporate bond mutual funds between February 2003 and December 2017. The fund flow shocks are computed as the residuals of the monthly fund flows after deducting the predicted flows from the past fund performance and the lagged flows. The Herfindahl–Hirschman Index is the sum of the squared share of the absolute monthly flow shocks across all active funds in the sample averaged within the year.

Year	Equity Funds (%)	Bond Funds (%)
1991	2.41	
1992	2.69	
1993	3.11	
1994	2.37	
1995	1.71	
1996	2.11	
1997	1.26	
1998	1.21	
1999	1.07	
2000	1.09	
2001	0.85	
2002	0.85	
2003	0.93	1.92
2004	0.74	1.84
2005	0.74	1.75
2006	0.56	2.19
2007	0.69	2.80
2008	0.52	2.12
2009	0.57	3.61
2010	0.56	1.96
2011	0.61	3.16
2012	0.68	2.74
2013	0.93	3.25
2014	0.76	3.15
2015	0.57	3.41
2016	0.51	2.98
2017	0.69	3.11

Table OA.8: Placebo Tests Using Index Funds in Cross-Sectional Asset Pricing

In this table, I construct the aggregate shocks to fund flows using index funds, and run Fama-MacBeth regressions of the bond returns on the instrumented fund flow betas with firm-month fixed effects and the other controls in equation (14) to estimate the risk prices for the fund flow betas and liquidity-adjusted market beta for the sample period between January 2008 and December 2017. The liquidity beta on flows, $\beta^{c,\hat{FLOW}}$, measures the co-movement of liquidity costs with the instrumented aggregate fund flow shocks, and the return beta on flows, $\beta^{r,\hat{FLOW}}$, measures the co-movement of returns with the instrumented aggregate fund flow shocks. The liquidity-adjusted market beta, β^{LMKT}, is the co-movement of net returns, i.e., returns minus liquidity costs, with the net market returns. To avoid look-ahead bias, I compute the instrumented fund flow betas and the liquidity-adjusted market beta using only the data available at the end of each year in the preceding 60-month rolling windows. Maturity is the number of years to maturity, Size is the log of the market capitalization, and Age is the number of years since issuance, as of the preceding month. For fixed effects, Seniority includes Junior, Junior Subordinate, Senior, Senior Subordinate, Senior Secured, and Subordinate bonds; Bond Type covers US Corporate Debentures, US Corporate MTN, and US Corporate MTN Zero; Credit Rating includes AAA, AA, A,..., C-rated bonds. GMM t-statistics are computed using the Newey and West (1987) method with two lags. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Liquidity Beta on Flows $\beta^{c,\hat{FLOW}}$	-0.050	-0.095	-0.061	-0.065	-0.061	-0.059
	(-0.59)	(-1.11)	(-0.69)	(-0.74)	(-0.69)	(-0.65)
Return Beta on Flows $\beta^{r,\hat{FLOW}}$	0.029	0.019	-0.006	-0.007	-0.005	0.012
	(0.77)	(0.46)	(-0.16)	(-0.17)	(-0.14)	(0.27)
Market β^{LMKT}	0.137**	0.161**	0.167**	0.171**	0.167**	0.173**
	(2.18)	(2.46)	(2.53)	(2.42)	(2.26)	(2.32)
Liquidity Costs	0.156***	0.150***	0.148***	0.148***	0.149***	0.148***
	(2.77)	(2.85)	(2.79)	(2.83)	(2.72)	(2.69)
Maturity	0.012*	0.010	0.010	0.010*	0.010*	0.010
	(1.93)	(1.61)	(1.66)	(1.66)	(1.68)	(1.62)
Size	0.022	0.028	0.029	0.030	0.030	0.032
	(0.87)	(1.16)	(1.19)	(1.23)	(1.23)	(1.30)
Age	0.001	0.001	0.000	0.000	0.000	-0.001
	(0.21)	(0.12)	(-0.02)	(-0.09)	(-0.09)	(-0.18)
GDP Growth	0.000*	0.001**	0.001**	0.001**	0.001**	0.001*
	(1.86)	(2.35)	(2.31)	(2.12)	(2.02)	(1.93)
Common Factor $\delta^{x,e,Size}$		-0.001**	-0.001**	-0.001**	-0.001**	-0.001**
		(-2.32)	(-2.32)	(-2.13)	(-2.04)	(-2.01)
Common Factor $\delta^{x,e,Active}$			0.002	0.001	0.002	0.002
			(0.66)	(0.39)	(0.51)	(0.67)
PC1 $\delta^{PCA,e,1}$				0.000	0.000	-0.001
				(0.14)	(-0.07)	(-0.34)
PC2 $\delta^{PCA,e,2}$					0.000	0.000
					(-0.21)	(-0.01)
PC3 $\delta^{PCA,e,3}$						0.004
						(0.89)
Firm-Month F. E.	Yes	Yes	Yes	Yes	Yes	Yes
Seniority-Month F. E.	Yes	Yes	Yes	Yes	Yes	Yes
Bond Type-Month F. E.	Yes	Yes	Yes	Yes	Yes	Yes
Credit-Month F. E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	231,191	231,191	231,191	231,191	231,191	231,191

Table OA.9: Cross-Sectional Asset Pricing Tests With Additional Lags

In this table, I include additional lags of fund flows and fund performance in the panel regression in equation (5), L={2, 3, 6}, and run Fama-MacBeth regressions of the bond returns on the instrumented fund flow betas with firm-month fixed effects and the other controls in equation (14) to estimate the risk prices for the fund flow betas and liquidity-adjusted market beta for the sample period between January 2008 and December 2017. Fund flow beta, $\beta^{F\hat{LOW}}$, is the return beta on flows minus the liquidity beta on flows, assuming a common risk price. The liquidity beta on flows, $\beta^{c,F\hat{LOW}}$, measures the co-movement of liquidity costs with the instrumented aggregate fund flow shocks, and the return beta on flows, $\beta^{r,F\hat{LOW}}$, measures the co-movement of returns with the instrumented aggregate fund flow shocks. The liquidity-adjusted market beta, β^{LMKT}, is the co-movement of net returns, i.e., returns minus liquidity costs, with the net market returns. To avoid look-ahead bias, I compute the instrumented fund flow betas and the liquidity-adjusted market beta using only the data available at the end of each year in the preceding 60-month rolling windows. Maturity is the number of years to maturity, Size is the log of the market capitalization, and Age is the number of years since issuance, as of the preceding month. For fixed effects, Seniority includes Junior, Junior Subordinate, Senior, Senior Subordinate, Senior Secured, and Subordinate bonds; Bond Type covers US Corporate Debentures, US Corporate MTN, and US Corporate MTN Zero; Credit Rating includes AAA, AA, A,..., C-rated bonds. GMM t-statistics are computed using the Newey and West (1987) method with two lags. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	$L=2$		$L=3$		$L=6$	
	(1)	(2)	(3)	(4)	(5)	(6)
Flow $\beta^{F\hat{LOW}}$	0.065***		0.067***		0.067**	
	(2.80)		(2.70)		(2.57)	
Liquidity Beta on Flows		-0.104**		-0.107***		-0.105***
		(-2.59)		(-2.62)		(-2.64)
Return Beta on Flows		0.049**		0.048**		0.05**
		(2.01)		(1.99)		(2.05)
Market β^{LMKT}	0.134**	0.166**	0.136**	0.161**	0.141**	0.162**
	(2.16)	(2.27)	(2.13)	(2.15)	(2.10)	(2.12)
Liquidity Costs	0.132**	0.142**	0.132**	0.141***	0.130***	0.140***
	(2.58)	(2.61)	(2.60)	(2.62)	(2.65)	(2.67)
Maturity	0.012*	0.010	0.012*	0.009	0.012*	0.009
	(1.98)	(1.65)	(1.98)	(1.61)	(1.97)	(1.61)
Size	0.019	0.032	0.019	0.032	0.019	0.032
	(0.78)	(1.35)	(0.79)	(1.34)	(0.76)	(1.33)
Age	0.000	-0.001	0.001	-0.001	0.001	-0.001
	(0.10)	(-0.24)	(0.12)	(-0.23)	(0.12)	(-0.24)
GDP Growth	0.003*	0.098*	0.003*	0.105*	0.003*	0.106*
	(1.82)	(1.78)	(1.82)	(1.92)	(1.91)	(1.93)
Common Factor $\delta^{x,e,Size}$		-0.001*		-0.001**		-0.001**
		(-1.88)		(-2.04)		(-2.04)
Common Factor $\delta^{x,e,Active}$		0.192		0.187		0.165
		(0.59)		(0.57)		(0.50)
PC1 $\delta^{PCA,e,1}$		-0.001		0.000		0.000
		(-0.45)		(-0.07)		(-0.04)
PC2 $\delta^{PCA,e,2}$		-0.001		-0.002		-0.002
		(-0.43)		(-0.74)		(-0.68)
PC3 $\delta^{PCA,e,3}$		0.003		0.004		0.004
		(0.89)		(1.11)		(1.21)
Firm-Month F. E.	Yes	Yes	Yes	Yes	Yes	Yes
Seniority-Month F. E.	Yes	Yes	Yes	Yes	Yes	Yes
Bond Type-Month F. E.	Yes	Yes	Yes	Yes	Yes	Yes
Credit-Month F. E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	210,267	210,267	190,795	190,795	174,294	174,294

Table OA.10: Double Sorts on Credit Ratings and Fund Flow Betas

I sort corporate bonds at the end of each year based on the credit ratings to form value-weighted tercile portfolios: 1) investment grade plus, with S&P ratings from AAA to A-; 2) investment grade minus, with S&P ratings from BBB+ to BBB-; and 3) speculative grade, with S&P ratings BB+ or lower. If the S&P rating is available, I use it; if not, I use Moody's, and if Moody's is also not available, I use the Fitch rating. Within each tercile, I construct tercile portfolios on the liquidity beta on flows, $\beta^{c,FLOW}$, in Panel A and the return beta on flows, $\beta^{r,FLOW}$, in Panel B. Both betas are computed in real-time using only the information available at the time of the sort to avoid look-ahead bias. Specifically, I estimate the aggregate shocks to fund flows in the expanding windows using all time-series from January 2003 to the time of sort, and then compute the historical fund flow betas in the preceding 60-month periods. I take the equal-average across the credit rating bins to obtain tercile portfolios on each fund flow beta. The monthly sample period is from January 2008 to December 2017. Newey-West t-statistics are in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Sorts on Liquidity Beta on Flows ($\beta^{c,FLOW}$)				
	1 (Low)	2	3 (High)	3-1
Panel A. Average Excess Returns				
$\mathbb{E}[r_{p,t} - r_{f,t}]$	0.690***	0.538***	0.443***	-0.247***
	(4.066)	(3.004)	(3.009)	(-3.862)
Panel B. CAPM				
CAPM α	0.709***	0.560***	0.462***	-0.247***
	(6.830)	(5.875)	(6.376)	(-3.877)

Panel B: Sorts on Return Beta on Flows ($\beta^{r,FLOW}$)				
	1 (Low)	2	3 (High)	3-1
Panel A. Average Excess Returns				
$\mathbb{E}[r_{p,t} - r_{f,t}]$	0.370***	0.515***	0.711***	0.340**
	(2.895)	(3.043)	(3.343)	(2.578)
Panel B. CAPM				
CAPM α	0.386***	0.535***	0.737***	0.351***
	(5.525)	(5.817)	(6.612)	(3.197)

Table OA.11: Portfolio Sorts on the Return Beta on Flows ($\beta^{r,FLOW}$)

This table presents portfolios on the return beta on flows, $\beta^{r,FLOW}$, which captures the co-movement of returns with the aggregate fund flow shocks. I sort the corporate bonds at the end of each year based on the return beta on flows, computed in the preceding 60-month rolling windows using only the information available at the time of sort to avoid look-ahead bias. I estimate the aggregate fund flow shocks in the expanding windows using all time series from January 2003 to the time of sort. 7-factor α is computed using the DEF and TERM of Fama and French (1993) and the five risk factors of Fama and French (2015). 9-factor α is computed by additionally including the two intermediary capital factors of He, Kelly, and Manela (2017). The monthly sample period is from January 2008 to December 2017. Newey-West t-statistics are in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	1 (Low)	2	3	4 (High)	4-1
Panel A. Average Excess Returns					
$\mathbb{E}[r_{p,t} - r_{f,t}]$	0.307***	0.419***	0.551***	0.710***	0.403**
	(3.245)	(3.594)	(3.325)	(2.885)	(2.238)
Panel B. CAPM					
CAPM α	0.319***	0.434***	0.572***	0.740***	0.420***
	(7.489)	(8.669)	(8.073)	(5.31)	(3.091)
Panel C. Fama-French 5 Factor Model + TERM + DEF					
7-factor α	0.282***	0.379***	0.533***	0.759***	0.478***
	(5.923)	(7.744)	(6.548)	(5.549)	(3.996)
Panel D. Fama-French 5 Factor + TERM + DEF + He, Kelly, and Manela 2 Factor Model					
9-factor α	0.276***	0.376***	0.522***	0.728***	0.451***
	(5.784)	(7.620)	(6.396)	(5.381)	(3.813)

Table OA.12: Double Sorts on Fund Flow Betas

In this table, I sort corporate bonds at the end of each year based on the return beta on flows to form quartile portfolios, and within each quartile I construct quartile portfolios on the liquidity beta on flows. The return (liquidity) beta on flows measures the co-movement of returns (liquidity costs) with the aggregate fund flow shocks. Both betas are computed in real-time using only the information available at the time of the sort to avoid look-ahead bias. Specifically, I estimate the aggregate shocks to fund flows in the expanding windows using all time-series from January 2003 to the time of sort, and then compute the historical fund flow betas in the preceding 60-month periods. I take equal-average across the return beta bins to obtain quartile portfolios on the liquidity beta on flows. 7-factor α is computed using the DEF and TERM of Fama and French (1993) and the five risk factors of Fama and French (2015). 9-factor α is computed by additionally including the two intermediary capital factors of He, Kelly, and Manela (2017). The monthly sample period is from January 2008 to December 2017. Newey-West t-statistics are in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	1 (Low)	2	3	4 (High)	4-1
Panel A. Average Excess Returns					
$\mathbb{E}[r_{p,t} - r_{f,t}]$	0.613***	0.521***	0.461***	0.483***	-0.131***
	(4.184)	(3.590)	(2.859)	(3.330)	(-2.757)
Panel B. CAPM					
CAPM α	0.631***	0.539***	0.482***	0.502***	-0.129***
	(7.476)	(7.912)	(6.448)	(7.785)	(-2.777)
Panel C. Fama-French 5 Factor Model + TERM + DEF					
7-factor α	0.620***	0.468***	0.495***	0.440***	-0.179***
	(6.205)	(5.885)	(6.059)	(5.917)	(-3.517)
Panel D. Fama-French 5 Factor + TERM + DEF + He, Kelly, and Manela 2 Factor Model					
9-factor α	0.593***	0.448***	0.493***	0.419***	-0.174***
	(6.053)	(5.724)	(5.972)	(5.767)	(-3.398)

Table OA.13: Double Sorts on Price Impact and Fund Flow Betas

I sort corporate bonds at the end of each year based on the flow-induced price impact to form value-weighted quartile portfolios, and within each quartile I construct quartile portfolios on the liquidity (return) beta on flows. The flow-induced price impact is computed as a weighted average of fund flows across all bond funds that hold the bond, using the funds' holdings as the weights, averaged over 20-quarter rolling windows in Panel A, and over various horizons of rolling windows in Panel B. The return (liquidity) beta on flows measures the co-movement of returns (liquidity costs) with the aggregate fund flow shocks. Both betas are computed in real-time using only the information available at the time of the sort to avoid look-ahead bias. Specifically, I estimate the aggregate shocks to fund flows in the expanding windows using all time-series from January 2003 to the time of sort, and then compute the historical fund flow betas in the preceding 60-month periods. I take the equal-average across the price-impact bins to obtain quartile flow-beta portfolios. The 9-factor α is computed using the DEF and TERM of Fama and French (1993), the five risk factors of Fama and French (2015), and the two intermediary capital factors of He, Kelly, and Manela (2017). The monthly sample period is from January 2008 to December 2017. Newey-West t-statistics are in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Controlling for Flow-Induced Price Impact

	1 (Low)	2	3	4 (High)	4-1
Liquidity Beta on Flows ($\beta^{c,FLOW}$)					
9-factor α	0.601***	0.517***	0.408***	0.348***	-0.253***
	(6.215)	(6.547)	(5.624)	(4.948)	(-5.042)
Return Beta on Flows ($\beta^{r,FLOW}$)					
9-factor α	0.274***	0.372***	0.532***	0.735***	0.460***
	(5.455)	(7.314)	(5.980)	(5.793)	(4.253)

Panel B: Controlling for Flow-Induced Price Impact Over Different Horizons

	3 years	1 year	2 quarters	1 quarter
Liquidity Beta on Flows ($\beta^{c,FLOW}$)				
9-factor α	-0.280***	-0.252***	-0.282***	-0.256***
	(-5.483)	(-5.295)	(-5.657)	(-5.050)
Return Beta on Flows ($\beta^{r,FLOW}$)				
9-factor α	0.466***	0.477***	0.434***	0.454***
	(4.048)	(4.207)	(4.024)	(4.028)

Table OA.14: Double Sorts on Standard Liquidity Risk and Liquidity Beta on Flows

I sort corporate bonds at the end of each year based on the standard liquidity risk in Acharya and Pedersen (2005) to form value-weighted quartile portfolios, and within each quartile I construct quartile portfolios on the liquidity beta on flows, $\beta^{c,FLOW}$. Specifically, in Panel A, I control for the sensitivities of liquidity costs to market returns by first sorting on the liquidity beta on market returns, $\beta^{c,MKT}$. In Panel B, I control for the sensitivities of liquidity costs to market liquidity costs by first sorting on the liquidity beta on market liquidity costs, $\beta^{c,LIQ}$. All liquidity betas are computed in the 60-month rolling windows prior to the time of sorts. I take equal-average across the standard-liquidity-risk bins to obtain quartile flow-beta portfolios. 7-factor α is computed using the DEF and TERM of Fama and French (1993) and the five risk factors of Fama and French (2015). 9-factor α is computed by additionally including the two intermediary capital factors of He, Kelly, and Manela (2017). The monthly sample period is from January 2008 to December 2017. Newey-West t-statistics are in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Controlling for Liquidity Beta on Market Returns ($\beta^{c,MKT}$)					
	1 (Low)	2	3	4 (High)	4-1
Panel A. Fama-French 5 Factor + TERM + DEF					
7-factor α	0.629***	0.511***	0.471***	0.367***	-0.261***
	(7.065)	(6.599)	(7.070)	(4.483)	(-6.187)
Panel B. Fama-French 5 Factor + TERM + DEF + He, Kelly, and Manela 2 Factor					
9-factor α	0.604***	0.487***	0.464***	0.353***	-0.252***
	(6.933)	(6.476)	(6.923)	(4.316)	(-6.026)
Panel B: Controlling for Liquidity Beta on Market Liquidity Costs ($\beta^{c,LIQ}$)					
	1 (Low)	2	3	4 (High)	4-1
Panel A. Fama-French 5 Factor + TERM + DEF					
7-factor α	0.623***	0.496***	0.487***	0.369***	-0.253***
	(6.874)	(7.242)	(7.231)	(4.137)	(-5.645)
Panel B. Fama-French 5 Factor + TERM + DEF + He, Kelly, and Manela 2 Factor					
9-factor α	0.598***	0.479***	0.476***	0.350***	-0.248
	(6.740)	(7.099)	(7.073)	(3.958)	(-5.497)

Table OA.15: Double Sorts on Sensitivity to Macro Uncertainty and Fund Flow Betas

I sort corporate bonds at the end of each year based on sensitivities to macro uncertainty to form value-weighted quartile portfolios, and within each quartile I construct quartile portfolios on the liquidity beta on flows (Panel A) and on the return beta on flows (Panel B). I estimate sensitivities to macro uncertainty by running regressions of bond liquidity costs in Panel A (bond returns in Panel B) on the AR(1) residual of the macroeconomic uncertainty of Jurado, Ludvigson, and Ng (2015) in the 60-month rolling windows prior to the time of sorts (i.e., the same windows used to compute the fund flow betas). I take equal-average across the macro-uncertainty bins to obtain quartile flow-beta portfolios. 7-factor α is computed using the DEF and TERM of Fama and French (1993) and the five risk factors of Fama and French (2015). 9-factor α is computed by additionally including the two intermediary capital factors of He, Kelly, and Manela (2017). The monthly sample period is from January 2008 to December 2017. Newey-West t-statistics are in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Sorts on Sensitivity to Macro Uncertainty and Liquidity Beta on Flows ($\beta^{c,FLOW}$)					
	1 (Low)	2	3	4 (High)	4-1
Panel A. Fama-French 5 Factor + TERM + DEF					
7-factor α	0.624***	0.519***	0.442***	0.403***	-0.222***
	(6.877)	(7.286)	(6.514)	(5.016)	(-5.489)
Panel B. Fama-French 5 Factor + TERM + DEF + He, Kelly, and Manela 2 Factor					
9-factor α	0.600***	0.499***	0.432***	0.387***	-0.214***
	(6.738)	(7.166)	(6.356)	(4.845)	(-5.322)
Panel B: Sorts on Sensitivity to Macro Uncertainty and Return Beta on Flows ($\beta^{r,FLOW}$)					
	1 (Low)	2	3	4 (High)	4-1
Panel A. Fama-French 5 Factor + TERM + DEF					
7-factor α	0.319***	0.422***	0.556***	0.794***	0.475***
	(5.280)	(7.343)	(6.100)	(6.090)	(4.042)
Panel B. Fama-French 5 Factor + TERM + DEF + He, Kelly, and Manela 2 Factor					
9-factor α	0.301***	0.422***	0.550***	0.766***	0.465***
	(5.118)	(7.270)	(5.983)	(5.927)	(3.924)

Table OA.16: Portfolio Tilts in Equity Mutual Funds

This table documents the cross-sectional relations between fund flow betas and portfolio tilts by active equity mutual funds from 1996.Q1 to 2017.Q4. In Columns (1) and (2), I run panel regressions of the portfolio weights of active equity mutual funds, $w_{k,t}^{MF}$, on the lagged fund flow beta, $\beta_{k,t-1}^{FLOW}$, controlling for the lagged liquidity-adjusted market beta, $\beta_{k,t-1}^{LMKT}$, with year-quarter fixed effects for stock k in quarter t. The fund flow beta, $\beta_{k,t-1}^{FLOW}$ is the return beta on flows, $\beta_{k,t-1}^{r,FLOW}$, minus the liquidity beta on flows, $\beta_{k,t-1}^{c,FLOW}$, assuming a common risk price. In Columns (2) and (4), I use the liquidity beta on flows, $\beta_{k,t-1}^{c,FLOW}$, and return beta on flows, $\beta_{k,t-1}^{r,FLOW}$, separately, relaxing the assumption of a common risk price. In Columns (3) and (4), I use the weight deviations from the market benchmark, $w_{k,t}^{MF}$ - $w_{k,t}^{M}$, as the dependent variable and run the panel regressions. All betas are computed quarterly in the preceding 60-month periods using the data available only up to the preceding quarter $t-1$ to avoid look-ahead bias. All variables are standardized to have means of zero and standard deviations of one. Standard errors are two-way clustered at stock and year-quarter level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
	$w_{k,t}^{MF}$		$w_{k,t}^{MF} - w_{k,t}^{M}$	
Flow $\beta_{k,t-1}^{FLOW}$	-0.092***		-0.050***	
	(-10.60)		(-8.94)	
Liquidity Beta on Flows $\beta_{k,t-1}^{c,FLOW}$		0.081***		0.058***
		(14.03)		(13.48)
Return Beta on Flows $\beta_{k,t-1}^{r,FLOW}$		-0.058***		-0.024***
		(-7.36)		(-4.82)
Market $\beta_{k,t-1}^{LMKT}$	-0.040***	-0.042***	-0.005	-0.007
	(-4.22)	(-4.48)	(-0.91)	(-1.24)
Year-Quarter F.E.	Yes	Yes	Yes	Yes
Observations	255,505	255,505	253,114	253,114

Table OA.17: Heterogeneity in the Aggregate Shocks to Fund Flows

This table shows the summary statistics of the aggregate shocks to fund flows in high-yield corporate bond mutual funds (HY Corp) and investment-grade corporate bond mutual funds (IG Corp) separately for the sample period between February 2003 and December 2017. I run a panel regression of the fund flows on the past fund performance and the lagged flows with time-fixed effects and compute the individual flow shocks as the sum of the time-fixed effects and the idiosyncratic residual flows. I then take the TNA-weighted average of the individual flow shocks and take the time-detrended residuals to obtain the aggregate shocks to fund flows.

	#Obs	Mean	Std.dev.	p10	p25	p50	p75	p90
HY Corp (%)	178	0.000	1.444	-1.791	-0.721	0.117	0.726	1.844
IG Corp (%)	178	0.000	0.677	-0.736	-0.343	0.023	0.321	0.819

Table OA.18: Heterogeneity in Portfolio Tilts: By Fund Active Share

This table documents the heterogeneity in corporate bond mutual funds' portfolio tilts across different levels of active share. Fund-quarter level active share is computed as the sum of the absolute differences between the portfolio weights and the market benchmark weights, divided by 2. Panel A documents the portfolio tilts by bond funds in the top tercile of the active share and Panel B by bond funds in the bottom tercile of the active share in the sample period between 2008.Q1 and 2017.Q4. In Columns (1) and (2), I run panel regressions of the portfolio weights of mutual funds, $w^{MF}_{k,b,t}$, on the lagged liquidity beta on flows, $\beta^{c,FLOW}_{k,b,t-1}$, controlling for the lagged return beta on flows, $\beta^{r,FLOW}_{k,b,t-1}$, and the liquidity-adjusted market beta, $\beta^{LMKT}_{k,b,t-1}$, for firm k, bond b, in quarter t. I include year-quarter fixed effects and credit rating group fixed effects in Columns (1) and (3) and firm-quarter fixed effects in Columns (2) and (4). In Columns (3) and (4), I use the weight deviations from the market benchmark, $w^{MF}_{k,b,t} - w^{M}_{k,b,t}$, as the dependent variable. All betas are computed quarterly in the preceding 60-month periods using the data available only up to the preceding quarter $t-1$ to avoid look-ahead bias. All variables are standardized to have means of zero and standard deviations of one. Standard errors are two-way clustered at firm and year-quarter level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: High-Active-Share Funds				
	(1)	(2)	(3)	(4)
	$w^{MF}_{k,b,t}$		$w^{MF}_{k,b,t} - w^{M}_{k,b,t}$	
Liquidity Beta on Flows $\beta^{c,FLOW}_{k,b,t-1}$	0.080***	0.056***	0.064***	0.045***
	(11.07)	(6.09)	(9.76)	(5.13)
Return Beta on Flows $\beta^{r,FLOW}_{k,b,t-1}$	-0.131***	-0.086**	-0.097***	-0.074**
	(-5.15)	(-2.62)	(-3.92)	(-2.36)
Market $\beta^{LMKT}_{k,b,t-1}$	0.131***	0.087***	0.095***	0.069***
	(6.17)	(4.30)	(4.95)	(3.53)
Year-Quarter F.E.	Yes	No	Yes	No
Credit Rating Group F.E.	Yes	No	Yes	No
Firm-Quarter F.E.	No	Yes	No	Yes
Observations	150,131	134,676	150,130	134,675

Panel B: Low-Active-Share Funds				
	(1)	(2)	(3)	(4)
	$w^{MF}_{k,b,t}$		$w^{MF}_{k,b,t} - w^{M}_{k,b,t}$	
Liquidity Beta on Flows $\beta^{c,FLOW}_{k,b,t-1}$	0.063***	0.049***	0.044***	0.039***
	(6.76)	(7.50)	(5.07)	(6.52)
Return Beta on Flows $\beta^{r,FLOW}_{k,b,t-1}$	-0.041	-0.036	-0.002	-0.022
	(-1.38)	(-0.97)	(-0.06)	(-0.63)
Market $\beta^{LMKT}_{k,b,t-1}$	0.090***	-0.012	0.052**	-0.035**
	(4.03)	(-0.71)	(2.52)	(-2.12)
Year-Quarter F.E.	Yes	No	Yes	No
Credit Rating Group F.E.	Yes	No	Yes	No
Firm-Quarter F.E.	No	Yes	No	Yes
Observations	137,919	123,781	137,918	123,780

C. Liquidity Costs Proportionate to Monthly Period

In Table 2, the coefficient estimates for the average portfolio-level liquidity costs, $\mathbb{E}[c]$, provide an economically relevant interpretation as follows. Recall that the normalized liquidity cost, c, is the average per-trade selling cost that investors pay only when they actually trade. So, for example, if investors hold securities on average for two months, the average per-trade selling cost would incur only half of the liquidity cost in the tests using monthly observations. To proxy the unknown holding period of security, I take two approaches as in Acharya and Pedersen (2005). The first approach is to model the inverse of the unknown holding period as a free parameter, τ, and have it estimated in the cross-sectional asset pricing test as in Table 2. In Table 2, τ is estimated to be between 0.21 to 0.42 across the asset pricing tests, which implies the average holding period of securities to be around 3-6 months.

The second approach is to calibrate the value of τ to be the average turnover from the sample. The turnover reveals the percentage of shares outstanding traded during the month. In my sample, the average turnover is 0.19 for all stocks used in the portfolio formation. This implies that it takes on average 5.26 $\approx$ 1/0.19 months to trade the total shares outstanding once. Hence, I divide the liquidity costs by the average holding period of 5.26 months to incur only the monthly portion of the liquidity costs

$$\mathbb{E}\left[r_{p,t} - r_{f,t} - \frac{1}{5.26}c_{p,t}\right] = \alpha + \lambda^{LMKT}\beta_p^{LMKT} + \lambda^{FLOW}\beta_p^{FLOW}, \tag{OA.30}$$

and the cross-sectional asset pricing results are reported in Table OA.1 of the online appendix.

With the economic restriction imposed on the holding period, the test results in Table OA.1 of the online appendix echo the performance of the two-factor model with the estimated holding period in Table 2. I show that fund flow beta exhibits significant explanatory power for the cross-sectional average excess returns, jointly and separately, with positive and significant risk prices and small individual pricing errors.

D. Construction of Test Portfolios

Since mutual funds face investment mandates that may restrict the set of assets they can invest in, I form test portfolios focusing on the assets that are likely to be traded or held by mutual funds.[76] For this, I examine the mutual fund ownership at the security level and find that the ownership is considerably higher in the NYSE stocks than in the Amex and the NASDAQ stocks.[77] Therefore, I focus on the NYSE stocks when constructing the test portfolios in Table 2. I form the portfolios at the end of each year, consistent with the finding in Chevalier and Ellison (1997) that mutual fund managers change their risk exposures in portfolio holdings at the end of December and September. After sorting at the end of each year, I hold the stocks in each portfolio over the next 12 months and take the value-weighted average across the stocks within each portfolio to compute the portfolio-level returns. I rebalance at the end of each year from 1996 to 2017. The year starts from 1996 because I compute the fund flow beta in the preceding 60-month rolling window. Similarly, for corporate bonds, I focus on bonds whose ownership is greater than the mean ownership (3.48%) of all bonds.

D.1. 100 Stock Portfolios

25 Liquidity and Fund Flow Beta Portfolios

I sort stocks into five portfolios based on the average Amihud (2002) illiquidity in the preceding year. I compute the annual illiquidity as the ratio of daily absolute returns over dollar volume averaged over the year.[78] Within each of the five illiquidity portfolios, I sort the stocks based on the real-time return beta on flows, $\beta_{k,t}^{r,FLOW}$, computed in the preceding 60-month rolling windows using only the information available at the time of sort. I estimate the aggregate shocks to fund flows in the expanding windows using all observations from January 1991 to the time of sort. I exclude stocks if the annual liquidity or the return beta on flows is missing.

[76]If the asset market is fully integrated and the market is complete, one stochastic discount factor should price all assets in the market. However, when the market is only partially integrated for reasons that are not directly studied in the model, for example, due to investment mandates that restrict the set of assets mutual funds can invest in, the test assets should also reflect this.

[77]The medians of the mutual fund ownership in the NYSE, Amex, and NASDAQ stocks are 9.26%, 1.92%, and 0.52%, respectively, in my sample. To compute the average mutual fund ownership at a security level, I take the sum of the holdings by all active mutual funds in each quarter, divide it by the quarter-end market capitalization of the security, and take the time-series average.

[78]The annual illiquidity measure is analogous to the monthly illiquidity measure, $ILLIQ_{k,t}$, as defined in (B.1). The only difference is that it is averaged over the year instead of the month. A sort by the normalized illiquidity results in the same portfolio construction since it adds/multiplies the identical coefficients to the Amihud (2002) illiquidity for all stocks to compute the normalized illiquidity.

25 Book-to-Market by Size Portfolios

I compute the size and book-to-market of each firm following Fama and French (1993) and Fama and French (2015). The size of a firm is measured as the market capitalization of the stock in December in the preceding year. I sort the stocks into five quintiles based on the size and within each size quintile I form five portfolios sorted on the book-to-market ratio. For a company with non-missing total assets and total liabilities in Compustat, I compute the book value of a firm as total assets minus total liabilities plus investment tax credit (if available) plus deferred taxes (if available) minus preferred stock (redemption value) from the preceding fiscal year. If the preferred stock (redemption value) is missing, I use preferred stock (liquidating value) or preferred/preference stock (capital), in order. If the book value is negative, I keep it as missing. I exclude stocks from the sorts if the size or book-to-market ratio for sorting is missing.

25 Profitability by Size Portfolios

I compute the operating profitability of each firm following Fama and French (2015), and size and book equity as in the 25 book-to-market by size portfolios. I sort the stocks into five quintiles based on size and within each size quintile I form five portfolios sorted on operating profitability. I compute operating profitability as the annual revenue minus the costs of goods sold, interest expense, and selling, general, and administrative expenses over the book equity of the preceding fiscal year. I require a security to have at least one of the following to compute operating profitability: costs of goods sold, interest expenses, or selling, general and administrative expenses in the preceding fiscal year. I exclude stocks from the sorts if the size or operating profitability for sorting is missing.

25 Investment by Size Portfolios

I compute the investment of each firm following Fama and French (2015), and the size as in the 25 book-to-market by size portfolios. I sort the stocks into five quintiles based on firm size and within each size quintile I form five portfolios sorted on investment. I compute the investment of a firm as the percentage change in the total assets from one year before the preceding fiscal year to the preceding fiscal year. I exclude stocks from the sorts if the size or investment for sorting is missing.

D.2. 50 Bond Portfolios

25 Liquidity and Fund Flow Beta Portfolios

Similar to the 25 liquidity and flow beta portfolios of stocks, I sort corporate bonds into five portfolios based on the average illiquidity in the preceding year. I compute the annual illiquidity as the ratio of monthly absolute returns over dollar volume averaged over the year. Within each of the five illiquidity portfolios, I sort the bonds based on the real-time return beta on flows computed in the preceding 60-month rolling windows using only the information available at the time of sort. I estimate the aggregate shocks to fund flows in the expanding windows using all observations from January 2003 to the time of sort. I exclude bonds if the annual liquidity or the return beta on flows is missing.

25 Credit Rating and Maturity Portfolios

In the spirit of Fama and French (1993), I sort bonds into five quintiles based on credit ratings and within each credit-rating quintile I form five portfolios sorted on maturity. I assign bonds with an AAA credit rating to quintile one, AA+, AA and AA- to quintile two, A+, A, and A- to quintile three, BBB+, BBB, and BBB- to quintile four, and the remaining ratings to quintile five. Within each credit-rating quintile, I assign bonds with maturities below 4 years to quintile one, 4-8 years to quintile two, 8-12 years to quintile three, 12-16 years to quintile four, and the remaining maturities to quintile five. I exclude bonds from the sorts if the credit rating or maturity for sorting is missing.

E. Construction of Corporate Bond Datasets Using TRACE and FISD

I pull the intraday bond transactions such as over-the-counter price and trading volume from FINRA's TRACE (Trade Reporting and Compliance Engine) and other bond characteristics, including bond type, coupon rate, credit rating, issue date, and maturity, from Mergent Fixed

Income Securities Database (FISD) between 2003 and 2017.

I follow the cleaning procedures in Dick-Nielsen (2009) and Dick-Nielsen (2014) to clear erroneous reports in the raw TRACE dataset.[79] I split the raw data into two subsamples: pre-2012 and post-2012, then remove cancellation, correction, and reversal for both subsamples. I then remove agency tradings and double countings.

I merge the TRACE dataset with the FISD dataset by CUSIP. I keep one credit rating for each bond following Dick-Nielsen et al. (2012): If the S&P rating is available, I use the S&P rating. If not, I keep the Moody's. If missing again, then the Fitch rating.

In selecting the sample, I adopt the filtering criteria commonly used in the asset pricing literature for corporate bonds (for e.g., Feldhütter, 2012; Goldberg and Nozawa, 2021). I keep fixed- and zero-coupon bonds removing bonds with floating coupon rates. I keep publicly-placed bonds in the U.S. removing privately placed bonds under the Rule 144a. I keep bond type that equals to U.S. Corporate Debentures (CDEB), U.S. Corporate Medium Term Note (CMTN), U.S. Corporate Medium Term Note Zero (CMTZ), and U.S. Corporate Paper (CP) removing bonds that are convertible, redeemable, putable, exchangeable, or have sinking fund. I remove bonds that have less than 6 months to maturity.

F. Fund Characteristics

Active Share

Following Gabaix and Koijen (2020), I compute the active share for fund j in quarter t as

$$AS_{j,t} = \frac{1}{2}\sum_{b\in j} |w^{MF}_{j,k,b,t} - w^{M}_{j,k,b,t}| \tag{OA.31}$$

where $w^{MF}_{j,k,b,t}$ is mutual fund j's portfolio weight and $w^{M}_{j,k,b,t}$ is the market benchmark portfolio weight for firm k, bond b, in quarter t. Similar to Gabaix and Koijen (2020), I compute the market portfolio, $w^{M}_{j,k,b,t}$, as the market-weighted portfolio using all bonds.

[79]I thank Qingyi (Freda) Song Drechsler for sharing the code that cleans the error in the TRACE Enhanced File.

G. Merging CRSP and Morningstar Mutual Fund Datasets

G.1. Raw CRSP Dataset Clean-Up

The CRSP Mutual Fund data comes directly from the WRDS server in SAS data format. I use *monthly_tna_ret_nav* as the main base file of the CRSP database that contains monthly fund returns, total net assets, and other fund characteristics from January 1991 to December 2017, and merge in other CRSP data files to prepare for merging with the Morningstar database. I delete observations when both monthly fund returns and total net assets are missing. Then, *crsp_fundno* uniquely identifies each share class, and a pair of *crsp_fundno* and month uniquely identifies each observation in the CRSP database. I clean up the raw CRSP database in the order presented in what follows.

Merging and Checking Tickers

I merge in historical tickers from *fund_hdr_hist*. If a ticker is missing, I then merge in another file, *fund_hdr*, which keeps the most recent ticker for each *crsp_fundno*, and I replace the missing tickers with the most recent ticker.

i. I check if a *crsp_fundno* has multiple tickers in a given month. There is no combination of *crsp_fundno* and month that has multiple tickers in a given month.

ii. I check if a *crsp_fundno* has multiple tickers over the entire sample period. In this case, I use the last (latest) ticker per *crsp_fundno* following Pastor, Stambaugh, and Taylor (2015).

iii. Conversely, I check if a ticker has multiple *crsp_fundno*s in a given month. In this case, I change the tickers to missing as in Pastor, Stambaugh, and Taylor (2015).

iv. I check if a ticker has multiple *crsp_fundno*s over the entire sample period. Tickers can get reused by another fund after the initial fund that had originally used it no longer exists. Thus, in general, each ticker can correspond to multiple *crsp_fundno*s. I take care of these tickers later in the merging algorithm.

Merging and Checking *CUSIPs*

Similary, I merge in historical *CUSIP*s from *fund_hdr_hist*. If a *CUSIP* is missing, I then merge in another file, *fund_hdr*, which keeps the most recent *CUSIP* for each *crsp_fundno*, and I replace the missing *CUSIP*s with the most recent *CUSIP*.

i. I check if a *crsp_fundno* has multiple *CUSIP*s in a given month. There is no combination of *crsp_fundno* and month that has multiple *CUSIP*s in a given month.

ii. I check if a *crsp_fundno* has multiple *CUSIP*s over the entire sample period. In this case, I use the last (latest) *CUSIP* per *crsp_fundno* following Pastor, Stambaugh, and Taylor (2015).

iii. Conversely, I check if a *CUSIP* has multiple *crsp_fundno*s in a given month. In this case, I change the *CUSIP*s to missing as in Pastor, Stambaugh, and Taylor (2015).

iv. I check if a *CUSIP* has multiple *crsp_fundno*s over the entire sample period. I take care of these *CUSIP*s later in the merging algorithm.

Checking Reversals

In order to prevent possible decimal-place mistakes, I check for extreme reversal patterns in the monthly total net assets following Pastor, Stambaugh, and Taylor (2015). I first compute the proportional changes in total net assets, dtna = (mtna-lag_tna)/lag_tna, and create a reversal variable, rev = (lead_tna-tna)/(tna-lag_tna). The reversal variable would be -1 if there is a decimal mistake, for example, 20m, 2m, 20m. If abs(dtna)>=0.5, -1.25<=rev<=-0.75, and lag_tna>=10mil, I change the total net assets to missing.

G.2. Raw Morningstar Dataset Clean-Up

I download the monthly fund returns, total net assets, and other fund characteristics from January 1991 to December 2017 from Morningstar Direct. I delete observations when both the fund returns and total net assets are missing. I exclude observations with share class type “Load Waived” because this share class type is open only to certain investors; they never have a *CUSIP*; tickers always end with“.lw’, which do not match CRSP; and the total net assets are all missing. Then, *secid* uniquely identifies each share class, and a pair of *secid* and month uniquely identifies each observation in the Morningstar database. I clean up the raw Morningstar database in the sequence presented below:

Merging and Checking Tickers

i. I check if a *secid* has multiple tickers in a given month. There is no combination of *secid* and month that has multiple tickers in a given month.

ii. I check if a *secid* has multiple tickers over the entire sample period. There is no *secid* that has time-varying tickers over the sample period: either a *secid* never has a ticker over the entire sample period or a *secid* has the same ticker over the entire sample period without any missing tickers. Therefore, there is no need to forward- and backward-fill the tickers as I did for the CRSP data.

iii. I check if a ticker has multiple *secid*s in a given month. In this case, I change the tickers to missing.

iv. I check if a ticker has multiple *secid*s over the entire sample period. I take care of these tickers later in the merging algorithm.

Merging and Checking *CUSIP*s

i. I check if a *secid* has multiple *CUSIP*s in a given month. There is no combination of *secid* and month that has different *CUSIP*s in the Morningstar database.

ii. I check if a *secid* has multiple *CUSIP*s over the entire sample period. There is no *secid* that has time-varying *CUSIP*s over the sample period: either a *secid* never has a *CUSIP* over the entire sample period or a *secid* has the same *CUSIP* over the entire sample period without any missing *CUSIP*s. Therefore, there is no need to forward- and backward-fill the *CUSIP*s as I did for the CRSP data.

iii. Conversely, I check if a *CUSIP* has multiple *secid*s in a given month. In this case, I change the *CUSIP*s to missing.

iv. I check if a *CUSIP* has multiple *secid*s over the entire sample period. I take care of these *CUSIP*s later in the merging algorithm.

Checking Reversals

As in the CRSP database, I check extreme reversal patterns in the monthly total net assets to prevent possible decimal-place mistakes. I first compute the proportional changes in total net assets, dtna = (mtna-lag_tna)/lag_tna, and create a reversal variable, rev = (lead_tna-tna)/(tna-lag_tna). If abs(dtna)>=0.5, -1.25<=rev<=-0.75, and lag_tna>=10mil, I change the total net assets to missing.

G.3. Complete Match Between CRSP and Morningstar

Following Pastor, Stambaugh, and Taylor (2015), I define a fund as completely matched if all share classes belonging to the fund are well matched. I identify a fund by *fundid* in Morningstar and a share class by *secid* and *crsp_fundno*. Each share class is well matched if and only if (i) the 60th percentile (over the available sample period) of the absolute value of the difference between the CRSP and Morningstar monthly returns is less than 5 basis points and (ii) the 60th percentile of the absolute value of the difference between the CRSP and Morningstar monthly total net assets is less than $100,000.

Well-Matched in the Merge by Ticker

In order to find a mapping between *secid* and *crsp_fundno* whose share classses are well matched, I first merge CRSP and Morningstar by ticker and month. After the merge, I identify the well matched share classes based on the differences in returns and total net assets between the two databases.

Well-Matched in the Merge by *CUSIP*

I then use *CUSIP* to merge CRSP and Morningstar to supplement the mapping between *secid* and *crsp_fundno* whose share classes are well matched. After the merge, I identify well matched share classes based on the differences in returns and total net assets between the two databases.

Complete Match

I combine the two mappings of well matched share classes that I separately identified from the ticker merge and the *CUSIP* merge. I then merge in the combined mapping to the Morningstar database to identify completely matched funds.

G.4. Merging CRSP and Morningstar

Following Pastor, Stambaugh, and Taylor (2015), I use CRSP as a base master file and merge in Morningstar funds. I keep only the completely matched funds in CRSP and in Morningstar.

G.5. Expense Ratios and Total Net Assets

I use the expense ratios from CRSP because it provides the exact start and end day for each expense ratio observation whereas Morningstar provides expense ratios over each fiscal year period. I change negative expense ratios to missing.

Total Net Assets

I change the total net assets of a share class in a given month to missing if either CRSP or Morningstar is missing the total net assets of the share class in the month. If the absolute difference of the total net assets between CRSP and Morningstar is greater than $100,000 and the difference is greater than 5% of the total net assets in CRSP, then I change the total net assets to missing. In all other cases, I use the total net assets value from CRSP.

G.6. Identifying Active Funds

Index Funds

I identify index funds by (i) the indicator variables provided in CRSP and Morningstar and (ii) searching for a keyword in fund names. CRSP provides *index_fund_flag* and Morningstar provides *enhanced_index* and *index_fund*. Index funds in CRSP have *index_fund_flag* values that are equal to B (index-based fund), D (pure index fund), or E (index fund enhanced). In Morningstar, if the value of *enhanced_index* or *index_fund* is "Yes," then the funds are identified as index funds. Lastly, I search for "index" in the CRSP fund names. Following this procedure, I create a variable *index_drop* and record 1 if it is an index fund.

Other Funds

Morningstar provides two variables that classify funds into different categories: *morningstar_category* and *primary_prospectus_bchmk*. I first forward- and backward-fill *morningstar_category* and *primary_prospectus_bchmk* within each *crsp_fundno*. I then search for a set of keywords in the two variables to classify the category of the funds. The sets of keywords are listed in Table OA.19. I first create the indicator variables *bond_funds*, *international_funds*, *sector_funds*, *target_funds*, *real_estate_funds*, and *other_non_equity* that record 1 if a fund contains any keywords listed in the column. I then create a variable *num_cat* that sums all the indicator variables.

G.7. Group Subclasses

I aggregate share classes to have a fund-level dataset. For each month, I sum the CRSP total net assets and Morningstar total net assets across subclasses; I take the TNA-weighted average of CRSP returns, Morningstar returns, CRSP expense ratios, and CRSP turnover ratios; and I take the maximum of *index_drop* and *num_cat*.

Table OA.19: Keywords for the Classification of Mutual Funds

The table shows the keywords that are used to classify mutual funds into bond funds, international funds, sector funds, target funds, real estate funds, and other non-equity funds, based on Morningstar Category (Panel A) and Primary Prospectus Benchmark (Panel B) provided by Morningstar for mutual funds from 1991 to 2017.

Panel A: Morningstar Category

Classification	Keywords
Bond funds	bank loan, convertibles, emerging markets bond, high yield bond, high yield muni, inflation-protected bond, intermediate government, intermediate-term bond, long government, long-term bond, multisector bond, muni california intermediate, muni california long, muni massachusetts, muni minnesota, muni national interm, muni national long, muni national short, muni new jersey, muni new york intermediate, muni new york long, muni ohio, muni pennsylvania, muni single state interm, muni single state long, muni single state short, nontraditional bond, short government, short-term bond, ultrashort bond, world bond
International funds	china region, diversified emerging mkts, diversified pacific/asia, emerging markets bond, europe stock, foreign large blend, foreign large growth, foreign large value, foreign small/mid blend, foreign small/mid growth, foreign small/mid value, global real estate, india equity, japan stock, latin america stock, pacific/asia ex-japan stk, world allocation, world bond, world stock
Sector funds	commodities broad basket, communications, consumer cyclical, consumer defensive, equity energy, equity precious metals, financial, health, industrials, miscellaneous sector, natural resources, technology, utilities
Target funds	target date 2000-2010, target date 2011-2015, target date 2016-2020, target date 2021-2025, target date 2026-2030, target date 2031-2035, target date 2036-2040, target date 2041-2045, target date 2046-2050, target date 2051+
Real estate funds	global real estate, real estate
Other non-equity	currency, long/short equity, managed futures, market neutral, multialternative, trading-inverse commodities, trading-inverse debt, trading-miscellaneous

Panel B: Primary Prospectus Benchmark

Bond funds	bond, treasury, govt, barclay, municipal, convertible, investment-grade, consumer price index, t-bill, dollor (as in the raw data)
International funds	ACWI, world, global, emerging markets, latin america, eafe, msci em
Sector funds	commodity
Target funds	target
Real estate funds	property, reit
Other non-equity	

References

Acharya, V. V. and L. H. Pedersen (2005). Asset pricing with liquidity risk. *Journal of Financial Economics 77*, 375–410.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets 5*, 31–56.

Chevalier, J. and G. Ellison (1997). Risk taking by mutual funds as a response to incentives. *Journal of Political Economy 105*, 1167–1200.

Dick-Nielsen, J. (2009). Liquidity biases in trace. *Journal of Fixed Income 19*, 43–55.

Dick-Nielsen, J. (2014). How to clean enhanced trace data. *Working Paper*.

Dick-Nielsen, J., P. Feldhütter, and D. Lando (2012). Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics 103*(3), 471–492.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*, 1–22.

Feldhütter, P. (2012). The same bond at different prices: Identifying search frictions and selling pressures. *Review of Financial Studies 25*(4), 1155–1206.

Frazzini, A. and L. H. Pedersen (2014). Betting against beta. *Journal of Financial Economics 111*, 1–25.

French, K. R. (2008). Presidential address: The cost of active investing. *Journal of Finance 63*, 1537–1573.

Gabaix, X. and R. S. J. Koijen (2020). In search of the origins of financial fluctuations: The inelastic markets hypothesis. *Working Paper*.

Goldberg, J. and Y. Nozawa (2021). Liquidity supply in the corporate bond market. *Journal of Finance 76*(2), 755–796.

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics 126*, 1–35.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics 47*, 13–37.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2015). Scale and skill in active management. *Journal of Financial Economics 116*, 23–45.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*, 425–442.

Stambaugh, R. F. (2014). Presidential address: Investment noise and trends. *Journal of Finance 69*, 1415–1453.