A Factor Framework for Cross-Sectional Price Impacts*

Yu An† Yinan Su‡ Chen Wang§

March 4, 2024

Abstract

We study how noise trading flows impact the cross-section of asset prices in a market where sophisticated investors enforce no-arbitrage. In our model, individual asset flows, aggregated at the factor level, drive fluctuations in factor risk premia, which in turn impact asset prices through beta pricing. This structure fits the reduced-form patterns of how each asset's flow impacts its own price and other assets' prices with only a few factor-level parameters. A model-implied trading strategy, designed to exploit the reversion of factor-level price impacts, delivers strong investment outcomes and improves the performance of a wide range of anomaly portfolios.

Keywords: cross-section, factor, flow, price impact, risk

JEL Codes: G11, G12

*We thank valuable comments and suggestions from Federico Bandi, Hank Bessembinder, Andrew Chen, Zhi Da, Darrell Duffie, Robin Greenwood, Zhiguo He, Ben Hébert, Shiyang Huang, Bryan Kelly, Serhiy Kozak, Jiacui Li, Dong Lou, Paolo Pasquariello, Nagpurnanand Prabhala, Alessandro Rebucci, Paul Schultz, Dongho Song, Yang Song, Zhaogang Song, Semih Üslü, Wei Wu, and discussants, Aditya Chaudhry, Thummim Cho, Badrinath Kottimukkalur, Xin Liu, Marcel Müller, Andrey Pankratov, Oleg Rytchkov, as well as conference and seminar participants at the Fed Board, Wolfe, MFA, Southern Methodist University, FMCG Conference, DC Junior Finance Conference, SoFiE Annual Conference, CICF, University of Macau, City University of Hong Kong, Chinese University of Hong Kong, SAFE Asset Pricing Workshop, and UT Dallas Finance Conference. All errors are our own.

†Carey Business School, Johns Hopkins University; yua@jhu.edu.

‡Carey Business School, Johns Hopkins University; ys@jhu.edu.

§Mendoza College of Business, University of Notre Dame; chen.wang@nd.edu.

1 Introduction

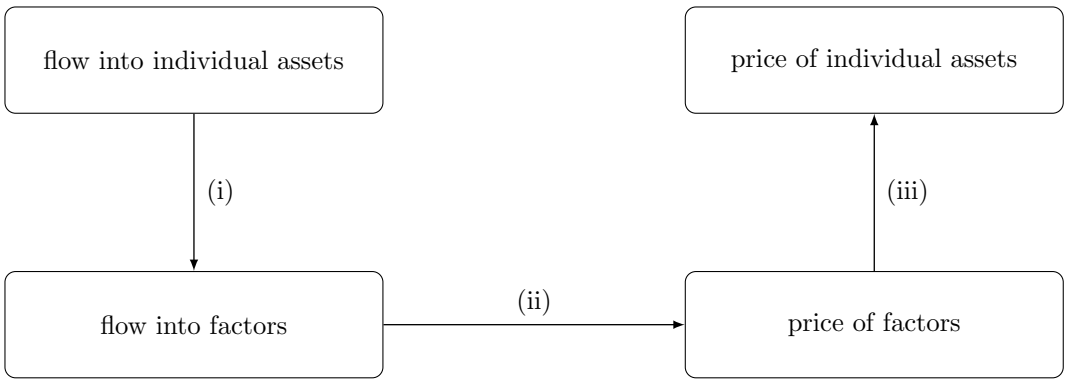
The interaction between noise traders and sophisticated investors is crucial in shaping asset prices. Empirical studies find that noise trading flows create large concurrent price impacts on individual assets and factor portfolios, and that sorting on asset-level flows generates anomalous future returns.[1] Theoretically, these phenomena are attributed to the limited risk-bearing capacity of sophisticated investors. As these investors absorb noisy flows by changing their risky asset holdings, asset prices adjust to compensate them for absorbing extra risks. Since these concurrent price impacts reflect the changes in risk premia, reversal strategies based on flows can generate abnormal returns.

Despite their limited risk-bearing capacity, sophisticated investors still enforce no-arbitrage pricing across assets—many empirical studies support that factor pricing governs the cross-section of expected returns (Fama and French, 1993 and Kozak, Nagel, and Santosh, 2018). In a market where noise traders and sophisticated investors interact, what structure should the price impacts of noisy flows take? Does capturing this structure better explain the reduced-form price impacts in the cross-section of assets?

We answer these questions with a framework in which noisy flows impact cross-sectional asset prices through risk factors. The framework is summarized in three steps (Figure 1). First, individual asset flows are aggregated to flows into a few factors. This aggregation accounts for the comovement of noisy flows—noise traders tend to buy and sell stocks with similar firm characteristics together. Second, factor-level aggregated flows change factor risk premia. The underlying mechanism is that market clearing dictates that what noise traders buy and sell is absorbed by sophisticated marginal investors. Consequently, factor flows change marginal investors' portfolio holdings and, in turn, their exposure to risk factors. Therefore, factor-level price impacts arise as the equilibrium compensation for marginal investors for absorbing risk. How much each factor's price changes in response to one unit of

[1] See, for example, Coval and Stafford (2007), Lou (2012), Chang, Hong, and Liskovich (2015), Koijen and Yogo (2019), Barber, Huang, Odean, and Schwarz (2022), and Gabaix and Koijen (2022).

Figure 1. Factor model of price impacts



Note: Noisy flows impact cross-sectional asset prices through risk factors.

flow-induced risk is the key parameter to estimate. This sensitivity, measured per unit of risk, differs from traditional price elasticity, which is per quantity of an asset. Third, individual asset prices respond to fluctuations in factor prices according to their risk exposures (beta). This step is canonical factor pricing with flow-driven conditional factor premium.

Our model integrates price impacts and factor pricing, but differs from either. To illustrate, consider two assets with identical exposures to risk factors, while only the first receives a unit of noisy flows. Our model implies that the prices of both assets should increase. Conversely, if the two risk exposures are of opposite signs, one asset's price increases and the other decreases. In other words, even with noise trading, beta pricing still holds in the cross-section of assets.[2] The underlying assumptions resonate with the Arbitrage Pricing Theory (APT): cross-asset arbitrage is relatively easy to enforce, whereas absorbing factor-level aggregate risk requires equilibrium compensation.[3] Unlike the APT, which estimates the unconditional price of risk (i.e., risk premium per unit of risk exposure), the price sensitivity in our model measures how flows drive fluctuations in the price of risk. Compared with the price impact framework, our model jointly estimates price impacts in the cross-section of assets, rather than estimating the price elasticity of each asset or factor portfolio

[2]Empirical support for this premise is found in Kozak, Nagel, and Santosh (2018). The distinctions between their findings and our work are discussed in detail in the literature review.

[3]Empirical studies documenting the existence of cross-asset arbitrage include Andrade, Chang, and Seasholes (2008) and Chaudhary, Fu, and Li (2023).

one-by-one in isolation (e.g., Gabaix and Koijen, 2022 and Li and Lin, 2022). Furthermore, the logit demand system in Koijen and Yogo (2019) features proportional substitution across different assets, which does not depend on the covariance between these assets' returns.[4] In our framework, however, risk plays a pivotal role.

We estimate the model with canonical factor and flow constructions. We construct stock-level flows using U.S. equity mutual fund flow-induced trading, following standard literature practices (Coval and Stafford, 2007 and Lou, 2012). According to the literature, such flows represent noise trading and have little information content.[5] The test assets are the Fama-French 5×5 size and book-to-market double-sorted portfolios, and the factors include market (MKT), size (SMB), and value (HML). The stock-level flows are then aggregated to the level of the test assets and factors. We use a monthly sample from 2000 to 2020.

The estimation has two stages. The first stage pins down the portfolios that the marginal investors are adjusting when flow shocks arise. That is, we want to figure out what assets marginal investors tend to buy and sell together. The answer lies in identifying a set of long-short portfolio weights, which represent the changes in marginal investors' positions in response to a one-dollar shift in factor-level flows. By market clearing, marginal investors absorb what noise traders buy and sell. Therefore, these portfolio weights can be recovered by the time-series regression coefficients of each test asset's flow on the factor flows.[6] The asset-level flows have strong common components, and the estimated portfolios capture the variation in marginal investors' positions well, with R^2 values ranging from 50% to 85%.[7]

[4]Specifically, proportional substitution means that if one asset becomes more attractive to an investor, the choice probability for all other assets reduces by the same percentage. See Section 3.3.2 of Train (2009) for a detailed discussion on substitution patterns under logit models.

[5]The empirical literature justifies the underlying assumptions that mutual fund investors are mainly uninformed retail investors (e.g., Ben-David, Li, Rossi, and Song, 2022b and Huang, Song, and Xiang, 2024). Furthermore, their purchases or sales of mutual fund shares prompt mutual funds to buy or sell individual stocks in proportion to lagged holdings (e.g., Lou, 2012).

[6]Our methodology echoes Alekseev, Giglio, Maingi, Selgrad, and Stroebel (2022), who estimate the sensitivities of mutual fund portfolio changes in response to heat shocks and use these sensitivities as weights to construct hedging portfolios against climate risk.

[7]Our model is agnostic about the specific drivers of these commonalities across assets, which may arise from sentiment shifts (Greenwood and Shleifer, 2014) or mutual fund ratings (Ben-David, Li, Rossi, and Song, 2022a).

The second stage fits the model to the panel of realized returns jointly. Specifically, each test asset's return is linked to the time-series fluctuation in factor flows and the cross-sectional dispersion in risk exposures toward the aforementioned portfolios. This joint estimation of the entire panel allows us to determine three structural parameters, one for each factor. Each parameter quantifies the price impact of one unit of flow-induced risk for that specific factor. Combined, we estimate a structural model of how the cross-section of asset prices changes as a function of the flows.

The empirical results support the hypothesis that noisy flows impact cross-sectional asset prices through risk factors, and deliver several new insights. First, our structural model accounts for the majority of flow-induced return variations in the cross-section. The model explains 7% of return variations for the 5×5 assets with only three parameters, each corresponding to the price sensitivity of one factor. In comparison, ignoring the cross-sectional relation and regressing each asset's return onto its flow in isolation yields an 8% R^2 with 25 reduced-form parameters. The reduced-form asset-level return-to-flow regression coefficients can be replicated by the model-implied counterparts. This is true for each asset's own price multiplier, as well as for cross multipliers constructed by regressing an asset's return on other assets' flows. Moreover, the price sensitivity parameter for the SMB and HML factors is approximately five times larger than that for the MKT factor. This reflects the heterogeneity in the arbitrage capital deployed to absorb flows along different risk dimensions. With back-of-the-envelope calculations, we estimate that active arbitrage capital amounts to \$150 billion for the MKT factor, and \$25 and \$15 billion for HML and SMB, respectively.[8]

Finally, we apply the estimated flow-price system to construct a trading strategy. In our model, noise trading creates temporary price impacts. The strategy capitalizes on the subsequent reversal of these temporary impacts by selling short when an inflow pushes up the price and vice versa. Since factors drive price fluctuations in our model, the strategy

[8]This finding aligns with the idea that investors may have distinct investment mandates for different style portfolios. For example, an insurance company focusing exclusively on large-cap stocks can absorb MKT flows but cannot elastically absorb SMB flows, as doing so would require trading small-cap stocks.

trades against the factor-level flows to time these factors and earn their premium realization. To achieve the optimal risk-return tradeoff among the factors, the strategy trades more aggressively against factors with a greater price sensitivity (between risk premia and flow). Therefore, our strategy differs from the conventional strategies that sort on stock-level flows irrespective of the cross-sectional factor structure. Our strategy earns an out-of-sample annualized Sharpe ratio of 0.5 by itself, exceeding conventional stock-level flow sorting or long- and short-term return reversal strategies in the same evaluation period.

More importantly, the strategy exploits the dynamic change in factor prices, which is a different source of return predictability from anomaly strategies in the "factor zoo" that hinge on unconditional factor premia. Based on the potential diversification benefit, we combine our flow-based strategy with each anomaly portfolio in a model-implied fashion and evaluate the gains in investment outcomes. Among the 154 anomaly portfolios in Fama and French (2015) and Jensen, Kelly, and Pedersen (2023), 140 or 91% experience a positive increase in their out-of-sample Sharpe ratio following the flow-based enhancement. The average change in the annualized Sharpe ratio is an increase of 0.3. The above results are robust across a series of alternative specifications.

In the rest of the paper, Section 2 reviews the related literature; Section 3 provides the theoretical foundation; Section 4 presents the empirical method; Section 5 reports the estimation results; Section 6 reports the model-implied trading strategies; and Section 7 concludes.

2 Related Literature

Our paper lies at the intersection of the noise trading literature and the factor pricing literature—each with its extensive history and recent advancements.[9] A closely related paper is Kozak, Nagel, and Santosh (2018), who argue that due to the arbitrage activity of

[9]For the noise trading literature, see Campbell and Kyle (1993), Daniel, Hirshleifer, and Subrahmanyam (2001), among others. For the factor pricing literature, see Fama and MacBeth (1973), Fama and French (1993, 2015), among others.

marginal investors, asset returns should exhibit a factor structure even in the presence of noise traders. We build upon their intuition, but focus on the price impacts of noise trading and show that such impacts are channeled through risk factors.

At one end of this literature intersection, many papers have examined the price impacts of individual assets or the cross-impacts between pairs of assets, estimating them *in isolation*. What sets our framework apart is that we *jointly* analyze the cross-section of price impacts within a structural model. A strand of literature documents the impact of noise trading on factor prices, which overlaps with step (ii) of our procedure (e.g., Teo and Woo, 2004; Ben-David, Li, Rossi, and Song, 2022a; Kang, Rouwenhorst, and Tang, 2022; Li, 2022; Li and Lin, 2022; Huang, Song, and Xiang, 2024; see Gabaix and Koijen, 2022 for a summary). Unlike these studies, our contribution is not merely to estimate factor-level price impacts, but to use them to explain asset-level price impacts through a structural model. Even within the scope of factor-level estimation, our measure improves over traditional price elasticity by capturing risk exposures and eliminating mathematical ambiguities that arise when applying the traditional metric to long-short portfolios.

Another segment of literature directly estimates cross-impacts between pairs of assets using reduced-form approaches (e.g., Boulatov, Hendershott, and Livdan, 2013; Pasquariello and Vega, 2015; Chaudhary, Fu, and Li, 2023). Their approaches often encounter the curse of dimensionality, because one needs to estimate N^2 pairs of cross-impacts for N assets, a number that further compounds when portfolios of these assets are considered. Our model addresses this issue by linking the cross-section of price impacts via risk factors. This approach simplifies the N^2 problem to estimating the price impacts of only a few factors. Through this, the cross-impacts between any pair of assets can then be inferred from the covariance structure of their flows and returns.

At the other end of the literature intersection, the canonical factor model builds on the premise that risk exposures determine asset prices. Our factor model contributes to the literature by showing that flows impact prices by altering risk exposures. Recently,

a growing body of literature explores new methods for forming asset-pricing factors using firm-level characteristics or trading signals (e.g., Harvey, Liu, and Zhu, 2016; Kozak, Nagel, and Santosh, 2018; Kelly, Pruitt, and Su, 2019; Giglio and Xiu, 2021; Kelly, Malamud, and Pedersen, 2023). Rather than using flows as characteristics to create new factors, we apply flows to time existing factors. From a practical standpoint, we provide a framework to exploit an alternative source of return predictability, distinct from the factor zoo literature.

A tangentially related strand of literature examines noise trader risk and limits of arbitrage (De Long, Shleifer, Summers, and Waldmann, 1990; Lee, Shleifer, and Thaler, 1991; Amihud and Mendelson, 1991; Shleifer and Vishny, 1997; Lamont and Thaler, 2003; Loewenstein and Willard, 2006). These papers allow differences in prices for assets with identical payoffs by treating noise trading as a new source of risk. Our framework still adheres to the law of one price, implying that two assets with identical payoffs should have the same price.

Finally, a large strand of literature investigates the asset pricing implications of commonality in trading flow and volume. Hasbrouck and Seppi (2001) document that flows into individual stocks exhibit a factor structure using NYSE's TAQ data. Dou, Kogan, and Wu (2022) and Kim (2020) demonstrate that the commonality in flows in and out of mutual funds is a risk factor for expected stock returns. Lo and Wang (2000) show that trading volume exhibits a factor structure, and Alvarez and Atkeson (2018) show that trading volume is a priced risk factor. Balasubramaniam, Campbell, Ramadorai, and Ranish (2021) find that Indian households' stock holdings exhibit a factor structure. Unlike these papers, in our setting, common flows into the cross-section of stocks are not risk factors per se, but create price impacts by changing factors' risk premia.

3 Theoretical Foundation

This section presents the theoretical foundation for the factor model of price impacts. The key assumption is that different groups of marginal investors accommodate noisy flows into specific factor portfolios. This is in contrast to the typical assumption that the same marginal

investor accommodates noisy flows into different factors. Our assumption is motivated by the empirical evidence that different portfolios have different price sensitivity to flow, as in Gabaix and Koijen (2022) and Li and Lin (2022). The assumption naturally generates such differential price sensitivity for different factors, which then leads to the factor model of price impacts through the law of one price (LOOP).

There are two periods, $t = 0$ and $t = 1$, with a gross risk-free rate R_F. The model features N assets. Each asset, denoted by $n = 1, 2, \ldots, N$, has a payoff X_n at time $t = 1$, with the payoff vector represented as[10] $\mathbf{X} = (X_1, X_2, \ldots, X_N)^\top$. We assume that the payoff $\mathbf{X}$ is jointly normally distributed and exhibits an exact K-factor structure spanned by the factors: $\mathbf{b}_1^\top \mathbf{X}, \mathbf{b}_2^\top \mathbf{X}, \ldots, \mathbf{b}_K^\top \mathbf{X}$. Specifically, $\mathbf{b}_k = (b_{1,k}, b_{2,k}, \ldots, b_{N,k})^\top$ denotes the portfolio shares of factor k, where one share of factor portfolio holds $b_{n,k}$ shares of asset n.

At time 0, noise traders buy q_k shares of factor k. We assume that the K factors have both uncorrelated payoffs and uncorrelated flows, i.e., $\text{cov}(\mathbf{b}_k^\top \mathbf{X}, \mathbf{b}_j^\top \mathbf{X}) = 0$ and $\text{cov}(q_k, q_j) = 0$ for $k \neq j$. This assumption is without loss of generality because factors with correlated payoffs and flows can always be rotated to be uncorrelated. See Appendix A.1 for technical details.

Let f_n denote the total flow into asset n. Because q_k shares of flow into factor k lead to $q_k b_{n,k}$ shares of flow into asset n, it follows that $f_n = \sum_{k=1}^K q_k b_{n,k}$, meaning that asset flow exhibits a K-factor structure. The following proposition summarizes this relationship.

PROPOSITION 1. *The relationship between asset flows and factor flows is*

$$\begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_N \end{pmatrix} = q_1 \begin{pmatrix} b_{1,1} \\ b_{2,1} \\ \cdots \\ b_{N,1} \end{pmatrix} + q_2 \begin{pmatrix} b_{1,2} \\ b_{2,2} \\ \cdots \\ b_{N,2} \end{pmatrix} + \cdots + q_K \begin{pmatrix} b_{1,K} \\ b_{2,K} \\ \cdots \\ b_{N,K} \end{pmatrix}. \tag{1}$$

Marginal investors on the other side of the market accommodate noisy flows into factors and determine asset prices based on their optimality conditions. We assume that for each

[10] We use bold font notation for matrices and vectors, and $\mathbf{A}^\top$ to denote the transpose of matrix $\mathbf{A}$.

factor k, a continuum of marginal investors with a total mass μ_k and a CARA risk-aversion parameter γ_k absorbs the flow q_k. As discussed, our model differs from traditional setups in that μ_k and γ_k can differ across factors. Intuitively, our model accounts for the potential segmentation of flow-absorbing investors for these factors.

We now determine the time-0 price of the k-th factor, denoted as a function $P_k(q_k)$ of the flow q_k. In equilibrium, the factor price $P_k(q_k)$ is determined such that each marginal investor finds it optimal to buy $-q_k/\mu_k$ shares of the factor, clearing the market. That is,

$$-q_k/\mu_k = \arg\max_y \mathbb{E}[-\exp(-\gamma_k((S_k/\mu_k + y)\mathbf{b}_k^\top \mathbf{X} - yR_F P_k(q_k))], \tag{2}$$

where y represents the change in each marginal investor's holding in factor k, and S_k is the total amount outstanding of factor k. The first-order condition of (2) implies that

$$P_k(q_k) = \lambda_k(q_k - S_k)\text{var}(\mathbf{b}_k^\top \mathbf{X}) + \frac{\mathbb{E}(\mathbf{b}_k^\top \mathbf{X})}{R_F}, \tag{3}$$

where $\lambda_k = \gamma_k/(\mu_k R_F)$. As one can see, λ_k determines the factor-level price sensitivity for each unit of factor risk $\text{var}(\mathbf{b}_k^\top \mathbf{X})$ induced by the factor flow q_k. Naturally, a greater risk aversion γ_k or a smaller mass μ_k of marginal investors for a given factor leads to a larger price sensitivity λ_k.

Next, we apply the LOOP to price individual assets relative to factors. That is, if two portfolios have the same payoff at time 1, they should have the same price at time 0. We denote the time-0 price of asset n as $P_n(\mathbf{f})$, where the vector of asset flows in Proposition 1 is denoted as $\mathbf{f} = (f_1, f_2, \ldots, f_N)^\top$. Denote the vector of asset prices as $\mathbf{P}(\mathbf{f}) = (P_1(\mathbf{f}), P_2(\mathbf{f}), \ldots, P_N(\mathbf{f}))^\top$. The LOOP implies that

$$\mathbf{P}(\mathbf{f}) = \sum_{k=1}^K \frac{\text{cov}(\mathbf{X}, \mathbf{b}_k^\top \mathbf{X})}{\text{var}(\mathbf{b}_k^\top \mathbf{X})} P_k(q_k) = \sum_{k=1}^K \left(\lambda_k(q_k - S_k)\text{cov}(\mathbf{X}, \mathbf{b}_k^\top \mathbf{X}) + \frac{\text{cov}(\mathbf{X}, \mathbf{b}_k^\top \mathbf{X})\mathbb{E}(\mathbf{b}_k^\top \mathbf{X})}{\text{var}(\mathbf{b}_k^\top \mathbf{X})R_F} \right). \tag{4}$$

Now, we simplify the asset prices in (4) to obtain an empirically implementable factor

model. Define *price impacts* as the time-0 percentage price change with and without flow $\mathbf{f}$,

$$\Delta\mathbf{p}(\mathbf{f}) = \left(\frac{P_1(\mathbf{f}) - P_1(\mathbf{0})}{P_1(\mathbf{0})}, \frac{P_2(\mathbf{f}) - P_2(\mathbf{0})}{P_2(\mathbf{0})}, \ldots, \frac{P_N(\mathbf{f}) - P_N(\mathbf{0})}{P_N(\mathbf{0})}\right)^\top. \tag{5}$$

Furthermore, define *fundamental returns* as the asset returns from time 0 to 1 in the absence of flow,

$$\mathbf{R}_0 = \left(\frac{X_1}{P_1(\mathbf{0})}, \frac{X_2}{P_2(\mathbf{0})}, \ldots, \frac{X_N}{P_N(\mathbf{0})}\right)^\top. \tag{6}$$

We can then simplify equation (4) as

$$\Delta\mathbf{p}(\mathbf{f}) = \sum_{k=1}^{K} \lambda_k q_k \text{cov}(\mathbf{R}_0, \mathbf{b}_k^\top \mathbf{X}). \tag{7}$$

So far, the asset flows $\mathbf{f}$, factor flows q_k, and portfolio shares $\mathbf{b}_k$ are measured in number of shares. To further simplify (7), we switch measurement to dollar values relative to the asset prices $\mathbf{P}(\mathbf{0})$, consistent with standard empirical practice in cross-sectional asset pricing (e.g., Fama-French portfolio weights). Although we still employ the same notation throughout the paper for $\mathbf{f}$, q_k, and $\mathbf{b}_k$, they are now all measured in dollars.[11] By implementing this unit change, we can further simplify equation (7).

PROPOSITION 2. *The factor model of price impacts in* (7) *simplifies to*

$$\Delta\mathbf{p}(\mathbf{f}) = \sum_{k=1}^{K} \lambda_k q_k \text{cov}(\mathbf{R}_0, \mathbf{b}_k^\top \mathbf{R}_0). \tag{8}$$

Equation (8) shows that the equilibrium price impacts $\Delta\mathbf{p}(\mathbf{f})$ for the cross-section of assets are influenced by the K factors' λ_k, the factor flows q_k, and the quantity of risk exposure to the factors, represented by $\text{cov}(\mathbf{R}_0, \mathbf{b}_k^\top \mathbf{R}_0)$.

Several remarks are in order here. First, because beta pricing holds under our model, the

[11] The conversion from the number of shares to dollar values proceeds as follows: the asset flow becomes $f_n \to f_n P_n(\mathbf{0})$, the flow factor becomes $q_k \to q_k P_k(0)$, and the portfolio weights become $b_n \to b_n P_n(\mathbf{0})/P_k(0)$. The portfolio relationship described in Proposition 1 remains unchanged after this conversion.

cross-section of price impacts $\Delta\mathbf{p}(\mathbf{f})$ explicitly depends on the risk exposure $\text{cov}(\mathbf{R}_0, \mathbf{b}_k^\top \mathbf{R}_0)$. Therefore, if two assets have the same risk exposures to some factor k, flows into factor k would cause the same amount of price impacts on these two assets. In other words, the cross-sectional variation in the price impacts $\Delta\mathbf{p}(\mathbf{f})$ arises from the cross-sectional variation in the risk exposures to the K factors. Because factor flows q_k are linked to individual asset flows f_n as in Proposition 1, our model generates a factor-based substitution pattern between individual assets, which we fit to the data in Section 5.4.

Second, when multiplying equation (8) with the factor's portfolio weights $\mathbf{b}_k$, the parameter λ_k becomes

$$\lambda_k = \frac{\mathbf{b}_k^\top \Delta\mathbf{p}(\mathbf{f})}{q_k \text{var}(\mathbf{b}_k^\top \mathbf{R}_0)}. \tag{9}$$

The denominator $q_k \text{var}(\mathbf{b}_k^\top \mathbf{R}_0)$ is the total amount of risk induced by the factor flow, while the numerator $\mathbf{b}_k^\top \Delta\mathbf{p}(\mathbf{f})$ is the factor-level price impact. Economically, λ_k quantifies the price impact of one unit of flow-induced risk for each factor k, reflecting the changing risk premium associated with factor k. Unlike traditional asset pricing, which estimates the price of risk (i.e., risk premium per unit of risk exposure), our λ_k measures how this traditional metric shifts in response to factor flow. Consequently, we term λ_k as *the price of flow-induced risk*. As shown in equation (3), our λ_k directly maps to theoretical parameters—the risk aversion γ_k and the mass μ_k of marginal investors who absorb factor-k flows.

Third, the literature uses price elasticity $(\Delta P/P)/(\Delta Q/Q)$ to measure how a 1% change in quantity impacts the price (see, for example, Gabaix and Koijen (2022) for a summary). In contrast, in our setting, λ_k in (9) measures how a single unit of risk induced by the flow impacts the price. There are two key distinctions:

- Economically: We measure risk exposures explicitly to understand the impact of flows on prices. This is rooted in the economic channel in which price impacts arise because marginal investors require compensation for absorbing the risk induced by flows.

- Technically: Most asset-pricing factors are long-short portfolios. For long-short port-

folios, the traditional measure $(\Delta P/P)/(\Delta Q/Q)$ is not a well-defined mathematical object. This is because the total quantity for long-short portfolios is zero ($Q = 0$), and division by zero is not allowed in mathematics.[12]

Simply put, the newly introduced measure, λ_k, serves a different purpose than the traditional price elasticity measure. It is designed to capture the impact of risk on price, specifically for flow-absorbing investors, and is mathematically well-defined for long-short portfolios where the traditional measure fails.

Fourth, when all factors have the same price of flow-induced risk λ_k, our factor model simplifies to the standard multi-asset price impact model in the literature. Specifically, if $\lambda_1 = \lambda_2 = \cdots = \lambda_k = \lambda$, then equations (1) and (8) imply that $\Delta\mathbf{p}(\mathbf{f}) = \lambda \text{var}(\mathbf{R}_0)\mathbf{f}$, a standard formula in the literature (refer to the survey article by Rostek and Yoon, 2023). Comparing this formula with our factor model (8), one observes that a meaningful factor structure for price impacts hinges on the premise that λ_k can vary across factors, a hypothesis that we confirm empirically for the Fama-French factors.

Finally, the assumptions required to derive the factor model (8) can be relaxed. This flexibility includes permitting approximate K-factor structures for payoffs and flows, rather than exact ones. It also allows for endogenous segmentation across the K factors. Additionally, the model can accommodate autocorrelated flows that arrive at different times in a dynamic equilibrium. An (2023) and An and Zheng (2023) consider these theoretical generalizations. In particular, even though our model is formulated in a static context, the cross-sectional factor structure in Proposition 2 remains the same when marginal investors also anticipate variations in future flows. Consequently, in a dynamic setting, the empirical estimation also follows the same procedure that we describe next.[13]

[12]The total quantity Q is computed similarly to the portfolio flow ΔQ in Proposition 1, which projects the asset-level amount outstanding onto a set of portfolio weights. In most empirical studies, the market factor is included and its portfolio weights align perfectly with the asset-level outstanding Q. Hence, as usually anticipated, the Q of the market factor equals the sum of all assets. Yet for all other portfolios, $Q = 0$.

[13]The only difference from the static setting lies in the theoretical interpretation of λ_k. In the static model, $\lambda_k = \gamma_k/(\mu_k R_F)$, with γ_k and μ_k representing the risk aversion and the mass of marginal investors of factor k. In the dynamic model, λ_k also depends on the autocorrelations of flows (An and Zheng, 2023).

4 Empirical Framework

Having established a theoretical basis for the factor model of price impacts, this section presents the general estimation procedure.

4.1 Constructing Portfolio Flows

Estimating the factor model of price impacts requires selecting N test assets and K factors, both of which can be portfolios of M underlying stocks.[14] Therefore, we start by describing how to aggregate stock-level flows into portfolio flows.

The observable data consists of a panel of flow $f_{m,t}^{\{\text{stock}\}}$ into stock m at time t. We aim to aggregate these stock-level flows into K portfolios, where $w_{m,k}$ represents the weight of stock m in portfolio k. The aggregation to the N test assets is similarly conducted. As shown by Proposition 1, flows first go into portfolios and then are allocated to individual stocks according to the portfolio weights. Thus, to aggregate stock-level flows to portfolio flows, one can perform a cross-section regression of stock flows $f_{m,t}^{\{\text{stock}\}}$ on portfolio weights $w_{m,1}, w_{m,2}, \ldots, w_{m,K}$ for each period t, with the regression coefficients being the portfolio flows $q_{1,t}, q_{2,t}, \ldots, q_{K,t}$.

To formulate this procedure, we collect the portfolio weight $w_{m,k}$ in the $M \times K$ matrix $\mathbf{W}$. By the portfolio flow theory presented in Proposition 1, the flow $q_{k,t}$ into portfolio k at time t is given by

$$\begin{pmatrix} q_{1,t} \\ q_{2,t} \\ \cdots \\ q_{K,t} \end{pmatrix} = (\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \begin{pmatrix} f_{1,t}^{\{\text{stock}\}} \\ f_{2,t}^{\{\text{stock}\}} \\ \cdots \\ f_{M,t}^{\{\text{stock}\}} \end{pmatrix}. \tag{10}$$

This aggregation method differs from the simple method of directly summing up stock-level

[14] Our framework is not limited to stocks. We use "stocks" because of our empirical application.

flows based on portfolio weights, which lacks a theoretical foundation.

4.2 Estimation Procedure

The estimation involves N test assets, K factors, and T periods. The data inputs consist of the return $r_{n,t}$ and flow $f_{n,t}$ of test asset n at time t, along with the flow $q_{k,t}$ of factor k at time t. These variables are defined for $n = 1, 2, \ldots, N$, $k = 1, 2, \ldots, K$, and $t = 1, 2, \ldots, T$. The construction of test asset and factor flows follows the approach detailed in Section 4.1. We then proceed in two steps. First, we estimate $\mathbf{b}_k = (b_{1,k}, b_{2,k}, \ldots, b_{N,k})^\top$, which is the portfolio weights of factor k in terms of the N test assets. Second, we estimate each factor's price of flow-induced risk λ_k.

In Section 3, our model requires the measurement of flows in dollar values for cross-sectional analysis, such that $\mathbf{b}_k$ can be interpreted as portfolio weights corresponding to each dollar invested. However, the model does not impose any intertemporal constraints on flow normalization. In our empirical application, we normalize flows using the total stock market capitalization from the preceding period. This approach accommodates the increasing total market capitalization observed in the data and economically implies that the risk-bearing capacity of marginal investors is proportional to the total market capitalization. In the regression analysis, we also remove the unconditional time-series mean of $r_{n,t}$, $f_{n,t}$, and $q_{k,t}$, and omit the intercept terms. This simplification does not affect the parameters of interest.

The data generating process assumes that the price impact, as described in our model (8), occurs repeatedly over time. In each period t, flows $f_{n,t}$ arrive, leading to price impacts across all assets. Each asset n also experiences a fundamental-driven return fluctuation in period t, denoted as $\xi_{n,t}$. Equation (8) in Proposition 2 assumes that the K factors exhibit uncorrelated flows and fundamental returns. However, the K factors in the data may not meet this condition, so a rotation is needed. The original factors are still denoted by $q_{k,t}$ (flows) and $\mathbf{b}_k$ (portfolio weights), while the rotated factors—those that have uncorrelated flows and fundamental returns—are denoted by $\tilde{q}_{k,t}$ and $\tilde{\mathbf{b}}_k$.

Given Proposition 2, the observed return $r_{n,t}$ of asset n in period t is modeled as follows

$$r_{n,t} = \sum_{k=1}^{K} \lambda_k \tilde{q}_{k,t} \text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t) + \xi_{n,t}. \tag{11}$$

Special attention should be paid to $\boldsymbol{\xi}_t = (\xi_{1,t}, \xi_{2,t}, \ldots, \xi_{N,t})^\top$, the fundamental returns of the N assets in period t. This term replaces $\mathbf{R}_0$ in equation (8), and now serves a dual purpose in equation (11). First, $\boldsymbol{\xi}_t$ represents the fundamental-return component of $r_{n,t}$. Second, it influences the price impact component by determining the quantity of risk exposure through the term $\text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t)$.

In the rest of this section, we present the two-stage procedure for estimating (11).

4.2.1 First-Stage Regression

The first stage of our regression analysis estimates $\mathbf{b}_k = (b_{1,k}, b_{2,k}, \ldots, b_{N,k})^\top$. These weights of the N test assets, $\mathbf{b}_k$, define the portfolio that the marginal investors are adjusting when factor-level flow shock $q_{k,t}$ arises. Intuitively, this stage recovers what test assets marginal investors tend to buy and sell together. By Proposition 1, we run time-series regressions of the flow into an asset $f_{n,t}$ on the contemporaneous factor flows $q_{k,t}$,

$$f_{n,t} = \sum_{k=1}^{K} b_{n,k} q_{k,t} + e_{n,t}. \tag{12}$$

All else equal, a one-dollar increase in factor-k flow results in an increase of $\$b_{n,k}$ in asset-n flow. By market clearing, a one-dollar increase in factor-k flow leads to a decrease of $\$b_{n,k}$ in marginal investors' holdings of asset n. Therefore, the flow sensitivities $\mathbf{b}_k = (b_{1,k}, b_{2,k}, \ldots, b_{N,k})^\top$ represent the changes in marginal investors' holdings of the N assets associated with a one-dollar factor-k flow.[15]

[15]The portfolio weights $b_{n,k}$ obtained in the first-stage regression should not be confused with the stock-level portfolio weights $w_{m,k}$ in (10). One set of weights pertains to the N test assets, while the other is associated with the M underlying stocks.

The residual $e_{n,t}$ in the first-stage regression (12) represents the asset-level idiosyncratic flows that are not explained by factor flows. Our factor model (11) does not use these idiosyncratic flows. The model's empirical success relies on selecting factors whose flows $q_{k,t}$ can account for a significant portion of common variations in asset flows $f_{n,t}$ in regression (12). Empirically, we indeed find a high regression R^2 using the Fama-French factors. Moreover, concerns about ignoring idiosyncratic flows are alleviated because each unit of idiosyncratic flows generates a smaller price impact compared to factor flows, as documented by Gabaix and Koijen (2022) and Li and Lin (2022). Appendix A.3 offers additional theoretical results on how idiosyncratic flows affect the mean-variance optimal strategy that capitalizes on these flows, which is constructed later in the paper.

4.2.2 Second-Stage Regression

The second stage implements a panel regression based on equation (11). Here, asset return $r_{n,t}$ is regressed on the product of factor flow $\tilde{q}_{k,t}$ and the quantity of risk exposure $\text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t)$. The regression estimates the price of flow-induced risk λ_k for each factor.

Several remarks are in order. First, as discussed, factors in equation (11) require rotation to ensure uncorrelated fundamental returns and flows. For technical details, see Appendix A.1. Second, the unobservable fundamental return $\boldsymbol{\xi}_t$ serves two roles: 1) as an input in (11) for calculating the quantity of risk $\text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t)$, and 2) as the regression residual. To address the unobservability, we employ an iterative procedure.[16] Initially, we set the fundamental return equal to the observed asset return (i.e., $\xi_{n,t} = r_{n,t}$) and carry out the second-stage regression. Following this, we use the regression residual, which corresponds to the model-implied fundamental return, as the new $\xi_{n,t}$. This procedure is repeated until $\xi_{n,t}$ reaches convergence.[17] Third, a consistent estimate for λ_k requires $\text{cov}(\tilde{q}_{k,t}, \xi_{n,t}) = 0$, meaning that the factor flow $\tilde{q}_{k,t}$ is uncorrelated with the concurrent fundamental return $\xi_{n,t}$.

[16] Using long-horizon returns as proxies for fundamental returns yields similar empirical results.

[17] In the empirical application, we obtain quick convergence within fewer than ten iterations under reasonable convergence criteria. This result is stable across different initial values.

We address this endogeneity in the empirical application using both OLS and IV methods.

5 Empirical Application

In this section, we apply the empirical framework in Section 4 to mutual fund flows and the Fama-French test assets and factors.

5.1 Data and Empirical Measures

We use the mutual fund flow-induced trading, as proposed by Coval and Stafford (2007) and Lou (2012). We employ the Fama-French 5×5 portfolios, which are sorted based on size and book-to-market equity ratios, as our test assets. We then measure the returns $r_{n,t}$ and flows $f_{n,t}$ of these portfolios. Additionally, we measure the flows $q_{k,t}$ into the Fama-French three factors. In what follows, we present the flow construction.

First, we compute mutual fund flows in dollar amounts following standard procedures. In particular, we retrieve monthly mutual fund returns and characteristics from the CRSP Survivorship-Bias-Free Mutual Fund database and quarterly holdings from the Thomson/Refinitiv Mutual Fund Holdings Data (S12). Our sample period is from 2000 through September 2020.[18] Our mutual fund sample comprises both active and passive mutual funds. We denote the total net assets (TNA) of mutual fund m at the end of month t as $\text{TNA}_{m,t}$, and the mutual fund's net-of-fee return in month t as $r_{m,t}^{\{\text{fund}\}}$. The mutual fund flow in dollar amount is defined as follows:

$$f_{m,t}^{\{\text{fund}\}} = \text{TNA}_{m,t} - \text{TNA}_{m,t-1}(1 + r_{m,t}^{\{\text{fund}\}}). \tag{13}$$

We cross-validate mutual funds' monthly returns and TNA obtained from the CRSP database

[18] The mutual fund industry witnessed significant growth and sustained inflows throughout the 1990s (Lou, 2012; Ben-David, Li, Rossi, and Song, 2022a). In the post-2000 era, the monthly flows of mutual funds have generally maintained relative stability, prompting us to start our sample period from 2000, aligning with Gabaix and Koijen (2022).

with corresponding data from Morningstar and Thomson/Refinitiv. In the process, we manually correct several data input inaccuracies. Details regarding this process are in Appendix C.

Second, we translate mutual fund flows into stock-level flows, using the established assumption in the literature that mutual funds buy or sell stocks in proportion to their prior holdings. Importantly, we employ the two-quarter-lagged mutual fund holding to transform mutual fund flow into stock-level flows. For instance, we use the fund holdings from Q4 of the preceding year for mutual fund flows occurring in April, May, and June. This lag is based on two considerations. Firstly, the flow-induced trading construction formulated by Lou (2012) utilizes one-quarter-lagged mutual fund holding. The lag is used because while uninformed retail investors primarily drive mutual fund flows, mutual fund managers may have private information.[19] The lag ensures that the constructed stock-level flows represent the non-discretionary component of mutual fund flows, which are induced by retail trading. Secondly, we impose a two-quarter lag to ensure the holding information is observable for the out-of-sample tradable strategy.[20] Our results remain robust if we alternatively follow the literature's one-quarter lag.

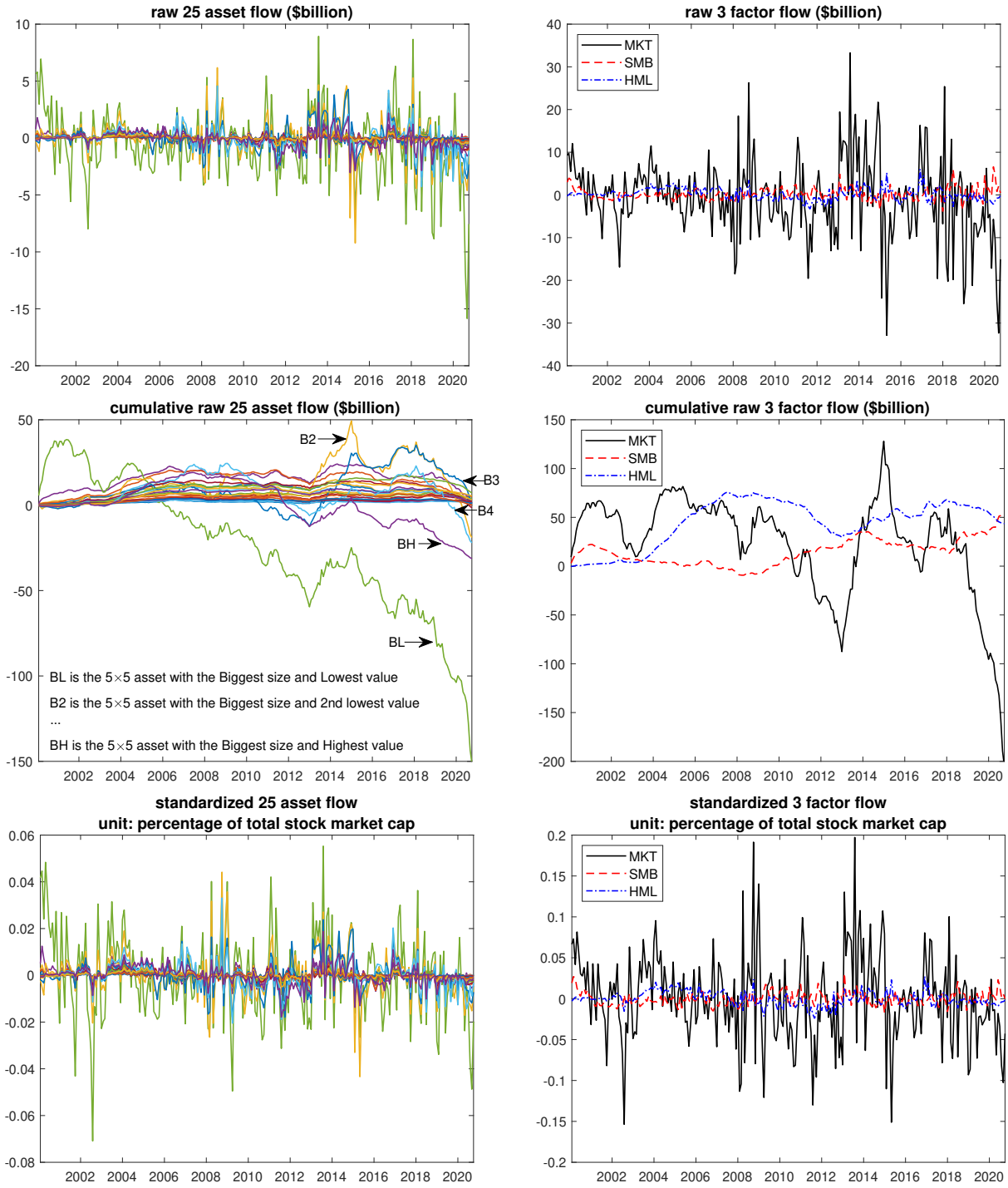
Third, we use stock-level flows and equation (10) to construct the Fama-French MKT, SMB, and HML flows $q_{k,t}$ and 5×5 test asset flows $f_{n,t}$. The top two panels of Figure 2 illustrate these monthly flows. The cumulative sum of test assets and factor flows are displayed in the two middle panels. Large stocks display substantial flow fluctuations due to their large market capitalization. Additionally, we notice a strong commonality among the flows of the 25 assets. The MKT flows display the most significant fluctuation, while the SMB and HML flows also show substantial variation.

As discussed in Section 4.2, we standardize the test asset and factor flows by dividing them by the total stock market capitalization from the previous month and subtracting the

[19] Frazzini and Lamont (2008) and Ben-David, Li, Rossi, and Song (2022b) provide evidence supporting this theory of uninformed mutual fund investors.

[20] The holding information is reported with a maximum statutory delay of 45 days (Christoffersen, Danesh, and Musto, 2015), which implies that Q1 holdings may not be observable in April. To remain conservative, we use the holding information from Q4 of the previous year for the flows in April, May, and June.

Figure 2. Time series of 25 test asset flows and three factor flows



Note: In the first two figures, we plot the monthly flows for the test assets and factors. In the subsequent two figures in the middle, we plot the cumulative sum of the flows for both test assets and factors. In the final two figures at the bottom, we standardize the test asset and factor flows by dividing them by the total stock market capitalization from the previous month and then subtracting the unconditional time-series mean. The sample period extends from January 2000 to September 2020.

unconditional time-series mean. The bottom two panels of Figure 2 plot the standardized test asset and factor flows, which are used in subsequent regressions. Notably, the MKT factor can experience inflows and outflows as large as 0.2% of the total stock market capitalization within a month.[21] Conversely, the SMB and HML flows display lower variations. The pairwise correlations between the flows of MKT and SMB, MKT and HML, and HML and SMB are 0.11, 0.25, and -0.11, respectively.

5.2 First-Stage Regression

Table 1 presents the results of the first-stage regression in equation (12),

$$f_{n,t} = b_{n,\text{MKT}} q_{\text{MKT},t} + b_{n,\text{SMB}} q_{\text{SMB},t} + b_{n,\text{HML}} q_{\text{HML},t} + e_{n,t}, \tag{14}$$

In this time-series regression, the flow $f_{n,t}$ of each asset n is regressed against the factor flows $q_{\text{MKT},t}$, $q_{\text{SMB},t}$, and $q_{\text{HML},t}$ to estimate $b_{n,\text{MKT}}$, $b_{n,\text{SMB}}$, and $b_{n,\text{HML}}$. The regression R^2 is provided in the upper-middle panel of Table 1, ranging from 50% for smaller companies to 85% for larger ones. These high R^2 values imply that factor flows account for a substantial proportion of common variations in asset flows. This finding supports our approach of using three factors to estimate the price impacts of the 5×5 assets.

As discussed in Section 4.2.1, flow sensitivities can be interpreted as portfolio weights. We now empirically investigate this relationship. First, the middle-left panel of Table 1 displays the MKT flow sensitivity $b_{n,\text{MKT}}$. For instance, a $b_{\text{BL,MKT}}$ value of 0.3252 implies that an increase of \$1 in the MKT flow results in a \$0.3252 increase in the BL asset flow. The top-left panel of Table 1 provides the market capitalization weight w_n of the 5×5 assets, calculated as the time-series average of the ratio of asset n's market capitalization to the total stock market capitalization. The value w_n serves as a proxy for the weights of the 5×5

[21] Flow into the MKT factor induced by mutual fund trading is less volatile than aggregate flow into mutual funds (which Appendix Figure A.1 shows) for two reasons. First, mutual funds do not invest 100% in stocks. Second, by our construction (10), when idiosyncratic flows are present, flow into the MKT factor is generally less volatile than the sum of flows into all stocks.

Table 1. First-stage regression: asset flows on factor flows

	market cap. weight w_n					regression R^2				
	Low	2	3	4	High	Low	2	3	4	High
Small	0.0042	0.0034	0.0040	0.0045	0.0044	48.03%	52.44%	52.61%	55.12%	52.13%
2	0.0081	0.0072	0.0071	0.0065	0.0042	59.38%	63.04%	58.81%	62.10%	63.67%
3	0.0142	0.0130	0.0107	0.0091	0.0060	60.50%	58.31%	62.55%	57.68%	62.68%
4	0.0371	0.0272	0.0190	0.0148	0.0112	57.79%	58.16%	54.72%	66.86%	67.10%
Big	0.3651	0.1790	0.1079	0.0844	0.0478	85.68%	85.68%	84.56%	87.44%	73.30%

	MKT flow sensitivity $b_{n,\text{MKT}}$					SMB flow sensitivity $b_{n,\text{SMB}}$					HML flow sensitivity $b_{n,\text{HML}}$				
	Low	2	3	4	High	Low	2	3	4	High	Low	2	3	4	High
Small	0.0022	0.0019	0.0027	0.0039	0.0033	0.0357	0.0298	0.0378	0.0357	0.0331	0.0258	0.0275	0.0377	0.0465	0.0464
2	0.0077	0.0073	0.0071	0.0073	0.0049	0.0844	0.0739	0.0649	0.0570	0.0378	0.0567	0.0697	0.0683	0.0681	0.0439
3	0.0139	0.0125	0.0096	0.0081	0.0046	0.1120	0.0848	0.0588	0.0455	0.0278	0.0895	0.1072	0.0948	0.0858	0.0591
4	0.0392	0.0252	0.0161	0.0143	0.0129	0.1153	0.0359	0.0003	0.0024	-0.0019	0.1045	0.1352	0.0976	0.0879	0.0855
Big	0.3252	0.1780	0.1169	0.0849	0.0575	0.0140	-0.2877	-0.2329	-0.2607	-0.1041	-0.3386	-0.1532	0.0654	0.2350	0.2007

sum of $b_{n,\text{MKT}}$ coefficients			sum of $b_{n,\text{SMB}}$ coefficients			sum of $b_{n,\text{HML}}$ coefficients		
all	positive	negative	all	positive	negative	all	positive	negative
0.97	0.97	0.00	0.10	0.99	-0.89	1.45	1.94	-0.49

	$t(b_{n,\text{MKT}})$					$t(b_{n,\text{SMB}})$					$t(b_{n,\text{HML}})$				
Small	4.64	4.34	4.80	6.60	4.74	8.65	9.40	7.73	6.07	5.37	8.18	9.85	9.96	12.44	11.20
2	7.64	8.48	8.13	8.63	8.31	10.40	8.66	7.63	6.85	6.75	9.96	14.25	13.65	12.81	14.46
3	8.21	8.34	8.12	6.83	6.24	10.55	7.20	5.65	5.11	5.00	10.32	12.56	16.26	13.29	15.54
4	10.38	9.39	8.44	10.30	9.29	5.47	2.52	0.02	0.29	-0.25	5.16	8.83	10.32	12.06	10.93
Big	21.37	22.34	25.74	22.49	9.98	0.23	-9.04	-9.11	-15.62	-5.92	-4.49	-3.61	2.79	12.17	6.66

Note: We run asset-by-asset time-series regressions (14) of asset flows on factor flows. The 5×5 assets are sorted based on size (small to big) and book-to-market equity (low to high). We report the regression R^2, point estimates and t-statistics for flow sensitivities $b_{n,\text{MKT}}$, $b_{n,\text{SMB}}$, and $b_{n,\text{HML}}$, as well as the market capitalization weight w_n of the 5×5 assets. To obtain w_n, we calculate the ratio of the market capitalization of asset n to the total stock market for each month. We then take a simple average of these values over time. The t-statistics are calculated from heteroskedasticity-robust standard errors.

test assets in the Fama-French MKT portfolio.

In line with the theory, for all assets n, we observe that the flow sensitivity $b_{n,\text{MKT}}$ is closely aligned with the Fama-French MKT weight w_n. All $b_{n,\text{MKT}}$ values are positive, and their sum is close to one. This finding aligns with the intuition that, on average, mutual funds hold the market portfolio. This finding also demonstrates that our new methods of constructing factor flows and interpreting flow sensitivities are consistent with the data.

Second, the middle-middle panel of Table 1 presents the SMB flow sensitivity $b_{n,\text{SMB}}$. We notice that $b_{n,\text{SMB}}$ resembles how Fama and French construct the small-minus-big factor. For instance, $b_{n,\text{SMB}}$ is positive for small companies, while for large companies, $b_{n,\text{SMB}}$ is negative. Furthermore, the sum of all positive $b_{n,\text{SMB}}$ values is 0.99, and the sum of all negative $b_{n,\text{SMB}}$ values is -0.89. The absolute value of both sums is close to one, once again aligning our empirical evidence with theoretical interpretation.

Third, the middle-right panel of Table 1 presents the HML flow sensitivity $b_{n,\text{HML}}$. We find that only the BL and B2 assets have negative flow sensitivities, while all other assets have positive flow sensitivities. The sum of positive $b_{n,\text{HML}}$ values is nearly 2, and the sum of negative $b_{n,\text{HML}}$ values is approximately -0.5. While these flow sensitivity estimates broadly align with how Fama and French construct the HML factor, there are significant numerical differences. What could explain this notable disparity? The key lies in the fact that the BL and B2 assets, representing large-growth companies, collectively account for over 50% of the total stock market capitalization. Therefore, the empirical evidence suggests that flows along the value direction are trading large-growth companies against all other companies. This finding substantially deviates from the Fama-French 2×3 construction.

5.3 Second-Stage Regression

Table 2 presents the results of the second-stage regression, as detailed in Section 4.2.2,

$$r_{n,t} = \sum_{k\in\{\text{MKT,SMB,HML}\}} \lambda_k \tilde{q}_{k,t} \text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t) + \xi_{n,t}. \tag{15}$$

Table 2. Second-stage regression: asset returns on factor flows × quantity of risk exposure

	total return OLS	intraday return OLS	intraday return IV
λ_{MKT}	9.54	5.99	6.76
	(14.48)	(12.07)	(2.83)
λ_{SMB}	109.93	64.42	151.56
	(9.84)	(6.19)	(3.49)
λ_{HML}	65.21	42.13	136.94
	(2.59)	(1.90)	(1.20)

Note: In this table, we run the second-stage regression of 5×5 asset returns on the product of factor flows and the quantity of risk to estimate the price of flow-induced risk. The unit of flow is expressed as a percentage of the total stock market capitalization, and the quantity of risk is expressed in terms of the annualized variance in returns. The first two columns display the OLS estimation results using total returns and intraday (open-to-close) returns. The third column outlines the IV estimation results using intraday returns, in which each factor flow is instrumented by the factor's concurrent overnight (close-to-open) return and the difference between one-month and half-year lagged flows. The figures in parentheses represent the t-statistics, computed using heteroskedasticity-robust standard errors.

Recall that one needs to rotate the MKT, SMB, and HML factors to obtain uncorrelated flows and fundamental returns. For rotated factors, $\tilde{q}_{k,t}$ represents the flows, and $\tilde{\mathbf{b}}_k$ denotes the portfolio weights.[22] Appendix Table A.1 presents the details of this rotation and shows that the rotated factors still resemble market, size, and value factors.

The first column of Table 2 presents the estimated price of flow-induced risk λ_k that represents the price impact of each factor in response to one unit of risk induced by the flow.[23] To interpret the estimated $\lambda_{\text{MKT}} = 9.54$, recall that the equilibrium condition (3) gives $\lambda_{\text{MKT}} = \gamma_{\text{MKT}}/(\mu_{\text{MKT}} R_F)$, where γ_{MKT} and μ_{MKT} denote the risk aversion and mass of the MKT factor's marginal investors, respectively, and R_F represents the gross risk-free rate. Assuming $\gamma_{\text{MKT}} \approx 3$ and $R_F \approx 1$, we obtain $\mu_{\text{MKT}} \approx 0.3$, meaning that 0.3% of the total stock market capitalization actively responds to flow as mean-variance optimizing marginal investors.[24] Though seemingly small, this figure aligns with Gabaix and Koijen (2022), who

[22]The rotation is to ensure a strict alignment between the empirical regression and the theoretical foundation in Section 3. Even without implementing this rotation, the second-stage regression results are similar.

[23]The t-statistics are highly significant and are calculated using heteroskedasticity-robust standard errors. We have also calculated standard errors clustered by year and found them to be smaller than the robust standard errors. To be conservative, we report t-statistics based on robust standard errors.

[24]Recall that the unit of flow is expressed as a percentage of the total stock market capitalization. Our choice of risk aversion $\gamma_{\text{MKT}} \approx 3$ is based on equation (9.6) in Cochrane (2009), combined with the fact that

find that the market is approximately 100 times less elastic than theoretical models suggest. In our context, this translates to less than 1% of the market actively responding to flows.

Performing the same calculation for the SMB and HML factors suggests that the market is even less elastic for these factors. Specifically, our estimates imply that 0.03% and 0.05% of the total stock market capitalization actively respond to SMB and HML flows, respectively. Given the current aggregate stock market capitalization of around $50 trillion, these percentages translate to approximately $15 billion and $25 billion in arbitrage capital, respectively. These empirical findings lend support to our theoretical assumption that different factors are met with different flow-absorbing capacities. One possible explanation for these results lies in the differing investment mandates. As we noted before, an insurance company that solely focuses on large-cap stocks could absorb MKT flows but would not be able to elastically absorb SMB flows, which would require trading in small-cap stocks.

Lastly, we address potential endogeneity issues concerning the estimates of λ_k. For an unbiased estimate, it is essential that the factor flow $\tilde{q}_{k,t}$ is uncorrelated with the fundamental return $\xi_{n,t}$. By constructing noisy flows $\tilde{q}_{k,t}$ using lagged fund holdings, we alleviate this endogeneity concern. This is because $\tilde{q}_{k,t}$ serves as a proxy for the portion of mutual fund flows mechanically driven by retail investors' buying and selling of mutual fund shares, rather than by the discretionary allocation of mutual fund managers. Retail investors are unlikely to be informed about $\xi_{n,t}$.

To address potential endogeneity concerns arising from mutual fund investors chasing fundamental returns, we perform additional robustness checks. We follow the methodology of Li (2022) and substitute asset return $r_{n,t}$ with intraday return $r_{n,t}^{\{\text{intraday}\}}$, which is the monthly aggregation of all open-to-close returns for each trading day.[25] The underlying premise is that intraday returns are more likely to reflect institutional trading, while overnight (close-to-open) returns are more likely to reflect retail trading (Lou, Polk, and Skouras, 2019, 2022;

the Sharpe ratio of the stock market is roughly 0.5 and the volatility is 0.16.

[25]Following Lou, Polk, and Skouras (2019), we construct the intraday and overnight returns on each date t as $r_t^{\{\text{intraday}\}} = \text{close}_t/\text{open}_t - 1$ and $r_t^{\{\text{overnight}\}} = (1 + r_t^{\{\text{close-to-close}\}})/(1 + r_t^{\{\text{intraday}\}}) - 1$. We source the price data from CRSP and adjust daily close-to-close returns for corporate actions such as stock splits.

Bogousslavsky and Muravyev, 2023). To isolate the price impacts driven by mutual fund flows, we utilize the intraday return $r_{n,t}^{\{\text{intraday}\}}$ as the dependent variable in regression (15). The regression results are reported in the second column of Table 2. The estimated λ_k is similar to, albeit smaller than, those in the first column.

Moreover, we supplement the intraday return OLS estimate with an IV strategy. The first instrument for $\tilde{q}_{k,t}$ is the factor's concurrent overnight return, represented by $\sum_{m=1}^{25} \tilde{b}_{m,k} r_{m,t}^{\{\text{overnight}\}}$. The IV relevance condition is met as factor flows are positively correlated with overnight returns, attributable to the return-chasing behavior of mutual fund investors. Furthermore, given that overnight and intraday returns are non-overlapping, the IV exclusion restriction is likely to be satisfied. The second instrument is $\tilde{q}_{k,t-1} - \tilde{q}_{k,t-6}$, which represents the difference between lagged flows over one month and half a year. The relevance of this instrument stems from the serial correlation observed in factor flows. The exclusion is because lagged flows are less likely to provide information about the fundamental return $\xi_{n,t}$. Importantly, using the difference in lagged flows, $\tilde{q}_{k,t-1} - \tilde{q}_{k,t-6}$, as the instrument, rather than the lagged flow $\tilde{q}_{k,t-1}$ itself, enables us to control for potential confounding channels, such as the possibility of lagged flow inducing future price reversion. The third column of Table 2 reports the IV results.[26] The point estimates largely align with those presented in the first two columns.

Additionally, to address concerns about serial correlation in flows biasing our λ_k estimates, we use the methodology of Lou (2012) to isolate unexpected flow components and reestimate Table 2. The results using unexpected flows, presented in Appendix Table A.3, align closely with our baseline findings in Table 2.

Up to this point, we have fully estimated the structural model of price impacts.

[26] The IV first-stage regression results are provided in Appendix Table A.2.

5.4 Evaluating Model Fit for Reduced-Form Price Impacts

We show that our model fits the reduced-form patterns of how each asset's flow impacts its own price and other assets' prices with only a few factor-level parameters.

5.4.1 Self Price Impact

We begin by evaluating the model's fit for the reduced-form self price impact, which is estimated by the time-series regression of each asset's return $r_{n,t}$ against its own flow $f_{n,t}$ normalized by the asset's market capitalization weight w_n,

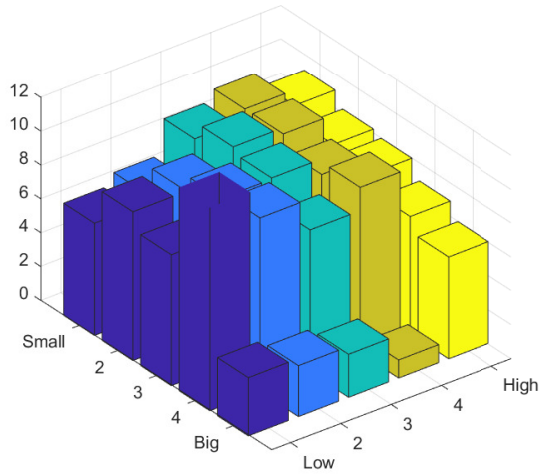
$$r_{n,t} = \eta_n(f_{n,t}/w_n) + \epsilon_{n,t}. \tag{16}$$

Here, $f_{n,t}/w_n$ quantifies the month-t flow into asset n as a proportion of the asset's market capitalization, so the estimated η aligns with the conventional price impact multiplier.

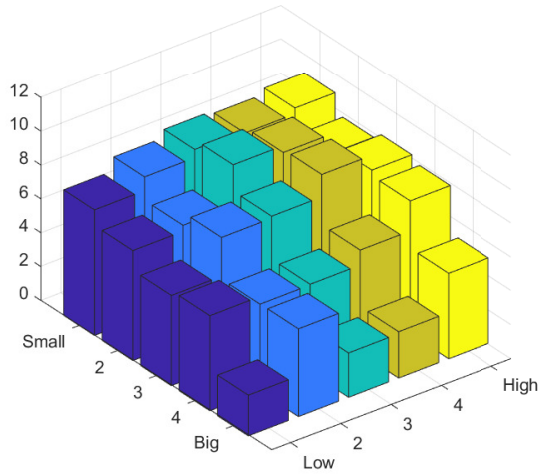
Figure 3 Panel A reports the reduced-form R^2 for each 5×5 asset from regression (16). Approximately 8% of the variation in realized returns can be attributed to an asset's own flow. In comparison, Panel B reports the R^2 from the factor-model regression (15).[27] The factor model explains about 7% of the return variation as price impact, as illustrated in Panel B and the summary statistics table below. This result means that the structural model fits the reduced-form patterns of how each asset's flow impacts its own price. Notably, our model estimates only 3 price sensitivity parameters, compared to the 25 in the reduced-form setup. This suggests that our estimates could be more robust in out-of-sample applications, as shown in the next Section 6. Panel C reports the reduced-form price multipliers η_n. In panel D, we replace the left-hand side of (16) with the model-fitted price impacts $\widehat{pi}_{n,t} := \sum_{k\in\{\text{MKT,SMB,HML}\}} \lambda_k \tilde{q}_{k,t} \text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t)$ and report the model-implied multipliers. The multipliers in both settings are around 10, further showing that the factor model

[27] For each n, $R_n^2 := 1 - \sum_t (r_{n,t} - \widehat{pi}_{n,t})^2 / \sum_t r_{n,t}^2$, where $\widehat{pi}_{n,t} := \sum_{k\in\{\text{MKT,SMB,HML}\}} \lambda_k \tilde{q}_{k,t} \text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t)$ is the model-fitted value of price impact in month t.

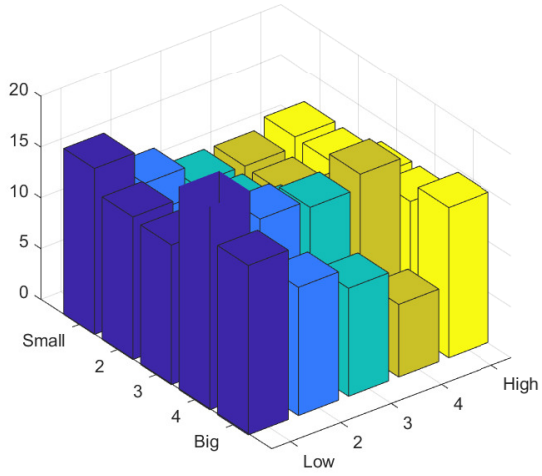
Figure 3. Regression R^2 and multipliers of reduced-form and model-implied self-impact



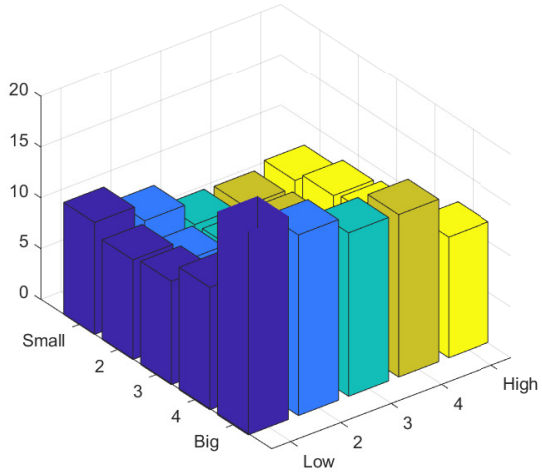
(A) reduced-form self R^2



(B) model-implied R^2



(C) reduced-form self multiplier



(D) model-implied self multiplier

	mean	std	P25	median	P75
reduced-form self R^2 (unit: %)	7.89	2.79	6.85	8.78	9.93
model-implied R^2 (unit: %)	6.67	2.07	5.19	7.41	8.30
reduced-form self multiplier	13.24	3.02	10.99	12.62	15.11
model-implied self multiplier	10.49	3.81	7.66	9.75	11.41

Note: Panel A reports the R^2 values from the reduced-form regression (16) for each test asset n. Panel B reports the R^2 values from the structured model (15). Panel C reports the reduced-form self-price multipliers η_n, estimated via (16). Panel D reports the model-implied self-price multipliers, estimated by replacing the left-hand side of (16) with the model-implied price impact. The table at the bottom provides the summary statistics for these R^2 values and multipliers.

captures the reduced-form patterns of price impacts in the cross-section.

5.4.2 Cross Price Impact

Next, we evaluate the model's fit for the reduced-form cross-asset price impact. Specifically, for each 5×5 asset n, we regress its return on the average flow into the adjacent assets on the 5×5 grid,

$$r_{n,t} = \phi_n \left(\sum_{m \text{ adjacent to } n} f_{m,t}/w_m \right) \Big/ (\text{number of } m \text{ adjacent to } n) + \epsilon_{n,t}, \tag{17}$$

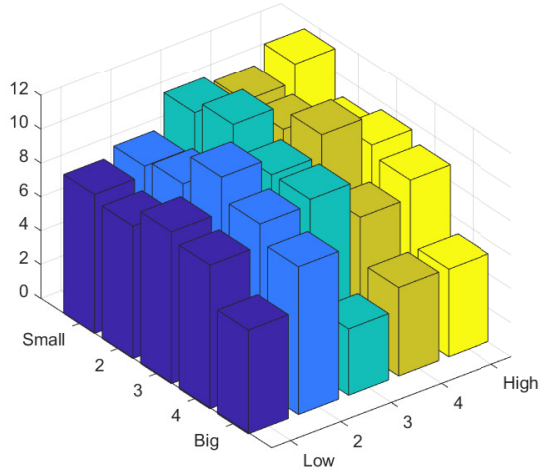
which examines the impact of these adjacent asset flows on each asset's return.

Figure 4 Panel A reports the reduced-form R^2 for each 5×5 asset from regression (17). Approximately 9% of the variation in realized returns can be attributed to adjacent flows. Panel B reports the R^2 from the factor-model regression (15), constructed in the same manner as in Figure 3 Panel B. This again demonstrates an alignment between the reduced-form and model-implied R^2 values.

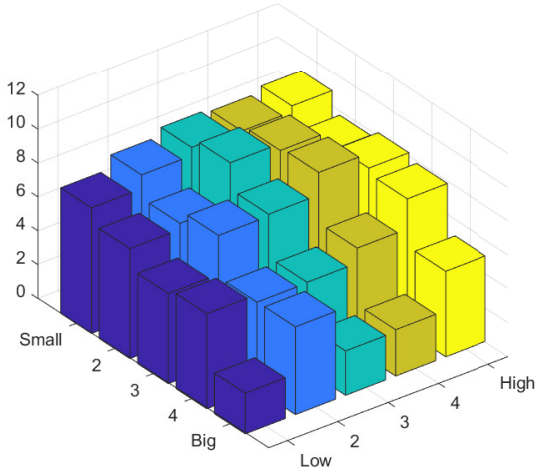
Panel C reports the reduced-form cross-price multipliers η_n, with an average value around 15. Interestingly, these cross-impact multipliers generally exceed the self-impact multipliers shown in Figure 3 Panel C. This means that cross multipliers are an important feature of the data that needs to be explained. In panel D, we replace the left-hand side of (17) with the model-fitted price impacts $\widehat{pi}_{n,t}$ and report the model-implied cross multipliers. Here too, we find an alignment between the reduced-form and model-implied cross multipliers.

In summary, our factor model fits both self and cross-asset price impacts of the 5×5 assets with only a few factor-level parameters. The evidence supports our initial hypothesis that noisy flows impact cross-sectional asset prices through risk factors.

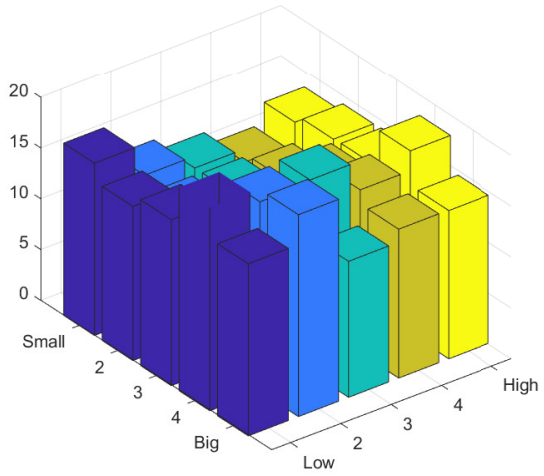
Figure 4. Regression R^2 and multipliers of reduced-form and model-implied cross-impact



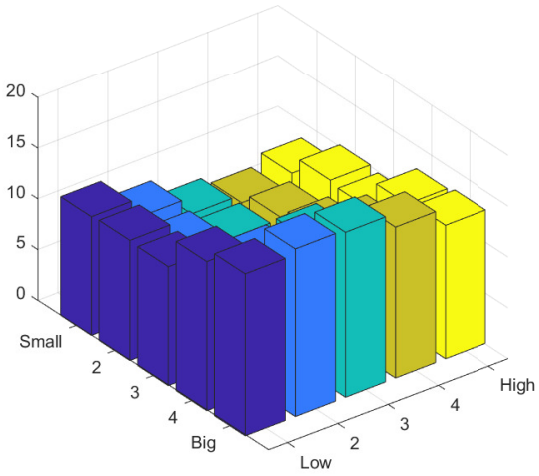
(A) reduced-form cross R^2



(B) model-implied R^2



(C) reduced-form cross multiplier



(D) model-implied cross multiplier

	mean	std	P25	median	P75
reduced-form cross R^2 (unit: %)	8.88	2.01	8.12	8.96	10.17
model-implied R^2 (unit: %)	6.67	2.07	5.19	7.41	8.30
reduced-form cross multiplier	15.18	2.47	13.33	14.58	16.93
model-implied cross multiplier	11.38	2.78	9.12	11.68	13.13

Note: Panel A reports the R^2 values from the reduced-form regression (17), where we regress each 5×5 asset's return on the average flow into the adjacent assets. Panel B reports the R^2 values from the structured model (15). Panel C reports the reduced-form cross-price multiplier ϕ_n, estimated via (17). Panel D reports the model-implied cross-price multipliers, estimated by replacing the left-hand side of (17) with the model-implied price impact. The table at the bottom provides the summary statistics for these R^2 values and multipliers.

6 Trading Strategy

In this section, we apply the estimated factor model to construct the model-implied optimal strategy to capitalize on the flow, and report its investment performance.

6.1 Strategy Construction

We begin by delving into the rationale behind our strategy. The underlying idea is that noisy flows temporarily distort prices from their fundamental values, implying that flow-induced price impacts should see long-term reversions. Intuitively, the mean-variance optimal reversion strategy should trade against the portfolio that has the maximum price impact per unit of risk. Consequently, we term this portfolio the Maximum-Price-Impact-Ratio (MPIR) portfolio. The associated strategy, which shorts this portfolio, is henceforth referred to as the MPIR strategy. Appendix A.2 presents the mathematical details of the mean-variance optimization of price impacts.

Given that our factor model (15) robustly characterizes the 5×5 assets' price impacts using only three factors, the model-implied MPIR strategy takes a simple form,

$$\sum_{k\in\{\text{MKT,SMB,HML}\}} -\lambda_k \tilde{q}_{k,t} \tilde{\mathbf{b}}_k. \tag{18}$$

The MPIR strategy shorts the factor portfolio $\tilde{\mathbf{b}}_k$ if there is a positive inflow $\tilde{q}_{k,t}$ and longs if there is an outflow. The dynamic strategy changes every month t depending on the factor flow $\tilde{q}_{k,t}$. The magnitude of the long/short position is proportional to the estimated λ_k.

At its core, the MPIR strategy times factors based on flow information. Three features of this strategy merit further discussion. First, although the MPIR strategy is formed using the 5×5 assets, its portfolio weights rely exclusively on the three factors' estimated risk compensation λ_k. Importantly, it does not use the 5×5 multipliers estimated in Figure 4. The rationale is that factor-level optimization, unlike asset-level optimization, delivers more

robust out-of-sample price reversion. This method is reminiscent of the traditional asset pricing approach, which employs factors to estimate the mean-variance optimal portfolio.

Second, the MPIR strategy differs from a simplistic strategy that indiscriminately trades against all flows, which can be represented by $\sum_{k\in\{\text{MKT,SMB,HML}\}} -\tilde{q}_{k,t}\tilde{\mathbf{b}}_k$. This distinction arises because the compensation λ_k for absorbing flow-induced risk varies across factors, as demonstrated in Section 5.3. Therefore, the optimal strategy trades more aggressively on factors that offer higher compensation per unit of risk.

Third, although the flow in month t distorts the price within the same month, the price might not revert entirely in month $t+1$, potentially taking longer. Consequently, to construct the portfolio for month $t+1$, we implement a staggered strategy that combines the portfolios (18) from the previous six months, from $t-5$ to t, using equal weights. Section 6.4 demonstrates the robustness of altering the initial and final months of the staggered strategy.

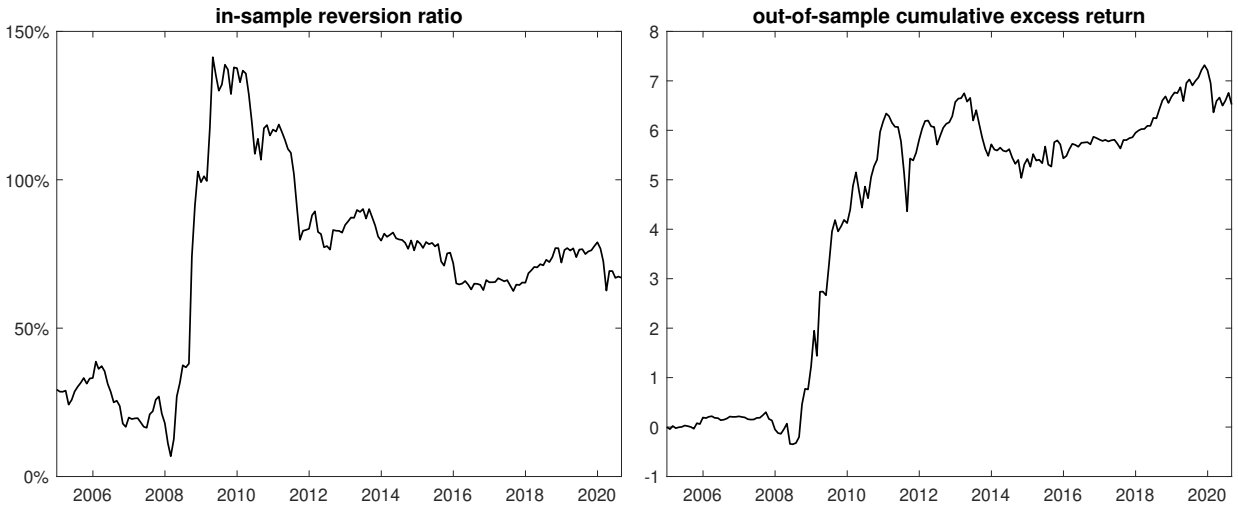
From a practical standpoint, the MPIR strategy is less susceptible to issues related to turnover and transaction costs. First, our strategy forms staggered portfolios based on the previous six months' flow, so only one-sixth of the positions need adjustment each month. Second, by its very design, our strategy provides liquidity to flows. Generally speaking, liquidity provision strategies are more likely to earn rather than pay bid-ask spreads.

6.2 Performance Evaluation

Our initial goal is to determine the effectiveness of the MPIR strategy within the sample. This utilizes expanding estimation windows, beginning from January 2000 and ending in December 2004, with the out-of-sample testing period extending from January 2005 to September 2020. Within each training window, the strategy is implemented, and the in-sample price reversion is computed. Subsequently, the ratio of the average reversion to the average model-implied price impact is calculated (see Appendix A.4 for details). Ideally, this reversion ratio should be 100%.

The left panel of Figure 5 presents the reversion ratio $\text{REV}^{\{t\}}$ corresponding to each

Figure 5. MPIR strategy's in-sample reversion and out-of-sample cumulative excess return



Note: In the left panel of the figure, we display $\text{REV}^{\{t\}}$, denoting the ratio of average one-month-forward reversion to the average model-implied price impact for the staggered MPIR strategy. This ratio is computed for each training window starting in January 2000 and ending in month t. The right panel depicts the cumulative excess returns of the MPIR strategy during the out-of-sample period, which spans from January 2005 to September 2020. For detailed strategy construction, refer to Appendix A.4.

training window that ends in month t. In most of the windows, this ratio falls below 100%, meaning that the model-implied price impact does not revert fully. Notably, the reversion exhibits considerable improvement following the financial crisis of 2008. This aligns with existing literature findings that after the crisis, marginal investors demonstrate reduced willingness to undertake substantial risk (Du, Tepper, and Verdelhan, 2018).

The fact that the in-sample reversion is less than 100% suggests a potential overestimation of the concurrent price impact in our model. To address this, for each training window, we normalize the estimated λ_k by $\text{REV}^{\{t\}}$ if $\text{REV}^{\{t\}}$ is less than 100%. These adjusted λ_k values are then employed to construct the out-of-sample MPIR strategy for the succeeding month $t+1$. Section 6.4 investigates the robustness of our approach when this normalization is not implemented.

The right panel of Figure 5 plots the cumulative excess returns of the MPIR strategy. For the out-of-sample period from January 2005 to September 2020, the strategy delivers an annualized Sharpe ratio of 0.49. The performance of the MPIR strategy markedly improves during the 2008 financial crisis, aligning with the fact that marginal investors demand

Table 3. The performance of MPIR strategy versus sorting-based strategies

	Return			Excess return			
Strategy	mean		std	mean		std	Sharpe ratio
MPIR	0.43	(2.01)	0.85	0.41	(1.94)	0.85	0.49
Trading-against-FIT	0.01	(0.30)	0.11	0.00	(-0.16)	0.11	-0.04
Trading-against-dollar FIT	0.00	(0.11)	0.07	-0.01	(-0.63)	0.07	-0.16
Short-term reversal	0.01	(0.34)	0.10	0.00	(-0.13)	0.10	-0.03
Long-term reversal	-0.04	(-1.60)	0.09	-0.05	(-2.14)	0.09	-0.54

Note: This table compares the performance of our MPIR strategy to sorting-based strategies. Specifically, we examine trading-against-FIT, trading-against-dollar FIT, short-term reversal, and long-term reversal. These are low-minus-high strategies sorted by 1) FIT (flow-induced trading), 2) dollar FIT, 3) one-month past return, and 4) 13 to 60-month prior returns, respectively. We report annualized statistics, including each strategy's sample mean (t-statistics in parentheses), the standard deviation for both raw and excess returns, and the Sharpe ratio. The sample period runs from January 2005 to September 2020.

higher risk compensation during periods of crisis, which tends to enhance the performance of reversion strategies (Nagel, 2012).

We further show that the MPIR strategy is distinct from and outperforms traditional sorting-based approaches. We examine two flow-based strategies: "trading-against-FIT" based on flow-induced trading as constructed in Lou (2012), and "trading-against-dollar FIT" based on flow-induced trading in dollar values. We also consider two widely-used return-based reversal strategies: "short-term reversal" from Jegadeesh (1990) based on the past one-month returns, and "long-term reversal" from De Bondt and Thaler (1985) based on the past 13 to 60-month returns. In each month, we sort stocks into quintiles using these metrics. We then create long-short portfolios by going long on the bottom quintile and short on the top quintile to capitalize on reversals. Table 3 reveals that the MPIR strategy outperforms these alternatives, both in Sharpe ratio and statistical significance.

6.3 MPIR Strategy Improves Existing Anomalies

The MPIR strategy exploits the dynamic change in factor prices, which is a different source of return predictability from strategies in the factor zoo that hinge on unconditional factor

premia. Based on potential diversification benefits, we hypothesize that our strategy should add on top of the investment performances of existing anomalies. That is, the Sharpe ratio of an existing anomaly should increase once we combine it with our strategy. Appendix A.2 provides a theoretical foundation for this assertion. This section substantiates it empirically.

We denote the MPIR strategy's excess return in month $t+1$ as $\tilde{r}^*_{t+1}$. We use the 154 anomaly portfolios from Jensen, Kelly, and Pedersen (2023), which include the Fama-French-Carhart six factors among a broad array of other firm characteristics-based anomaly portfolios.[28] We denote the excess return of anomaly portfolio j in month $t+1$ as $r_{j,t+1}$. The excess return of the combined portfolio j in month $t+1$ is defined as

$$r^*_{j,t+1} := r_{j,t+1} + w_{j,t}\tilde{r}^*_{t+1}, \tag{19}$$

where $w_{j,t}$ denotes the mean-variance optimal mixing ratio, which is estimated in-sample (refer to Appendix A.4 for detailed formula).

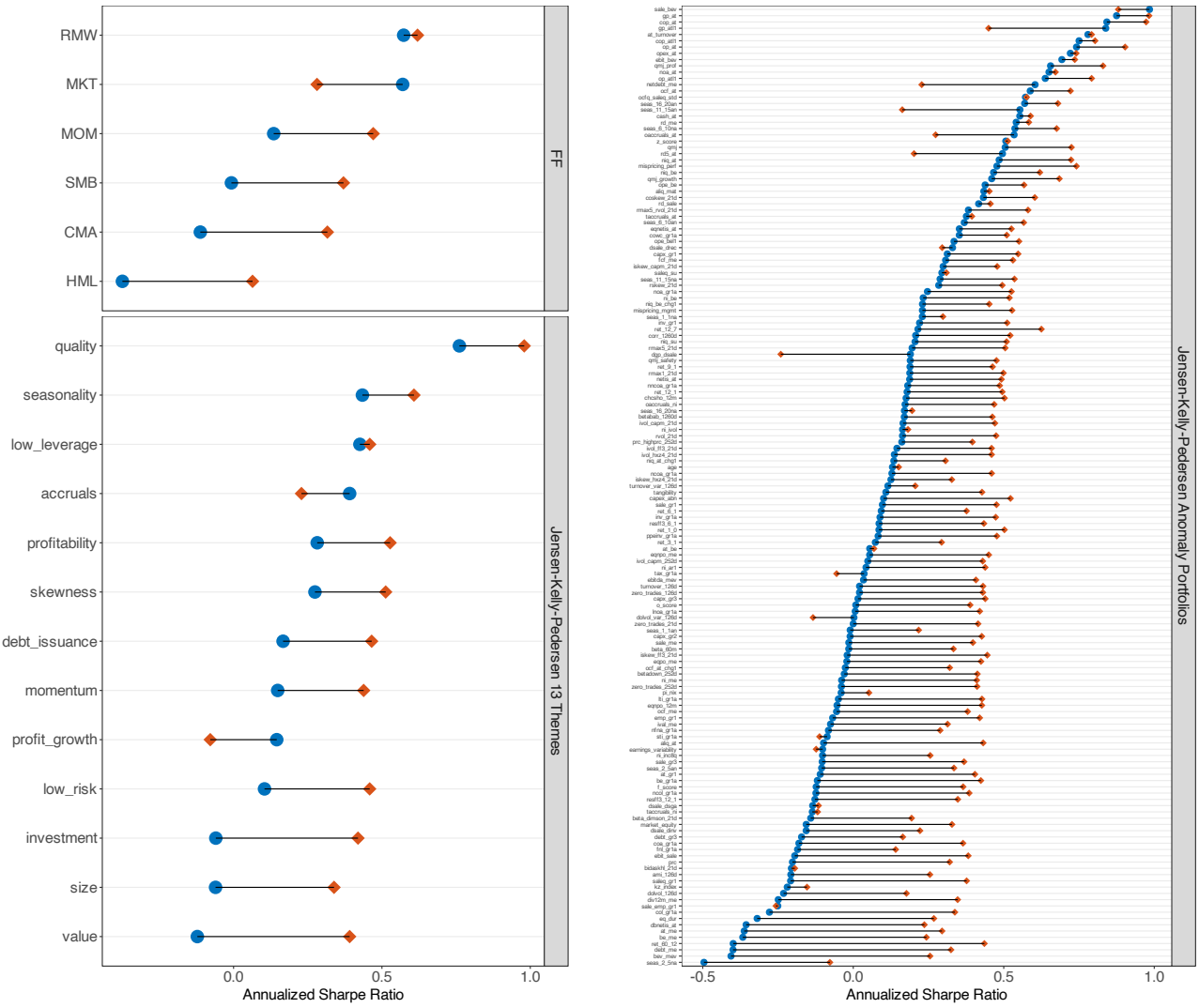
Figure 6 illustrates the improvement in the Sharpe ratio. The diagram is divided into three panels, showcasing the Fama-French-Carhart six factors, the Jensen-Kelly-Pedersen 13 themes, and individual anomaly portfolios, respectively. It is evident that the Sharpe ratios of the combined portfolios (represented by red diamonds) exceed those of the original anomaly portfolios (represented by blue dots) across the spectrum. Out of the 154 portfolios, 140, or 91%, exhibit a positive increase in the Sharpe ratio.[29] The average change in the Sharpe ratio is 0.26, and the median change is 0.30. This empirical evidence shows that our MPIR strategy improves the investment performance of existing anomalies.[30]

[28] We incorporate the 153 anomaly portfolios from Jensen, Kelly, and Pedersen (2023), along with the market excess return.

[29] The decrease in the Sharpe ratio for a small number of anomalies can be attributed to the weight $w_{j,t}$ becoming notably large during periods when the MPIR return $\tilde{r}^*_{t+1}$ is negative. This situation arises because $w_{j,t}$ is divided by the Sharpe ratio of anomaly j up to month t, in order to ensure proper risk weighting as described in equation (A.33) in the appendix. Consequently, if the historical Sharpe ratio is near zero, the weight $w_{j,t}$ can increase substantially.

[30] Appendix Figure A.2 presents a placebo test substituting $\tilde{r}^*_{t+1}$ from the MPIR strategy in equation (19) with short- and long-term reversal strategies. We find that the investment gains above existing anomalies vanish. This placebo test confirms that the enhanced performance over existing anomalies is not simply a mechanical effect of strategy combination.

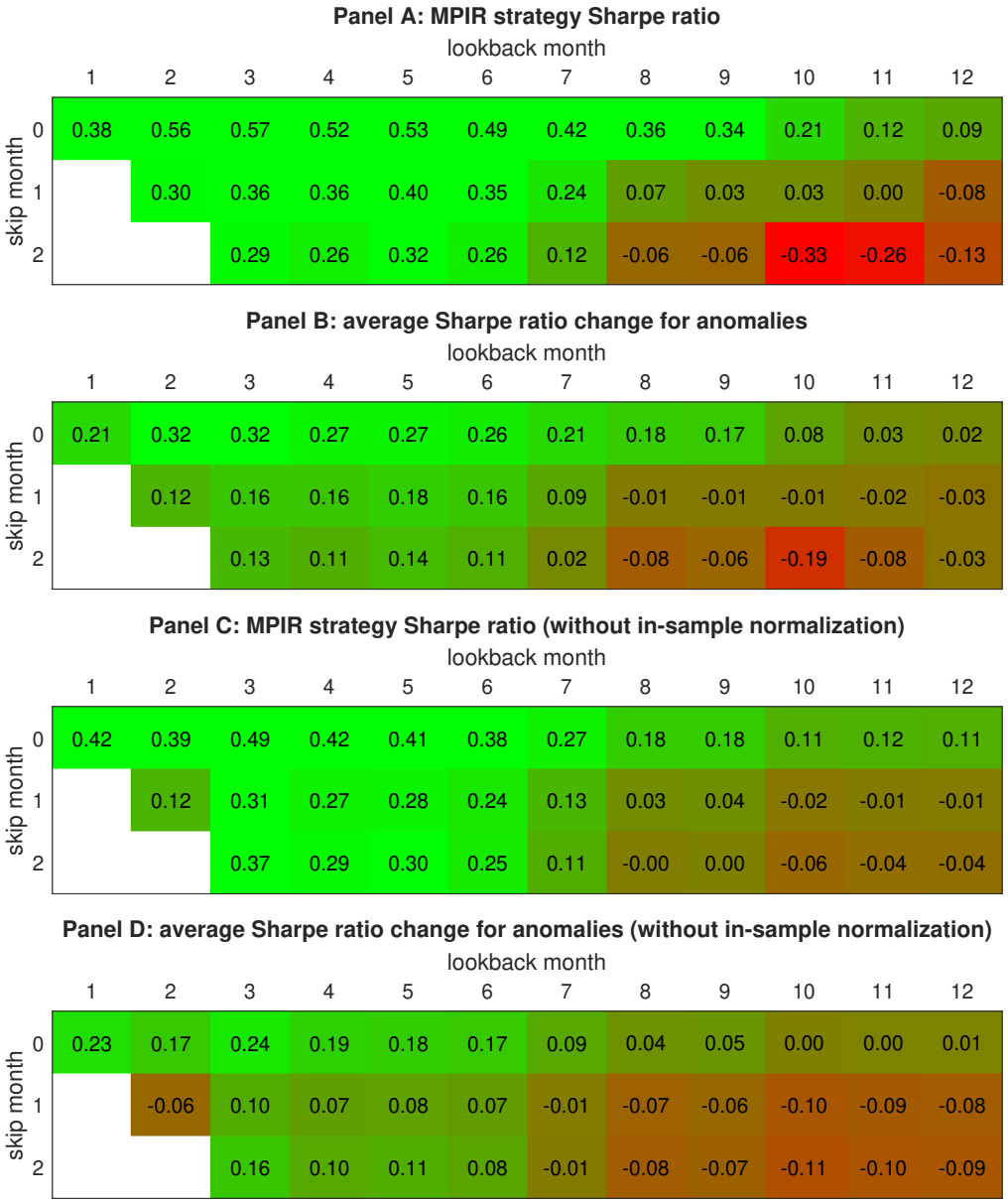
Figure 6. MPIR strategy increases out-of-sample Sharpe ratio of anomaly portfolios



	number of anomalies	mean	std	P25	median	P75
Sharpe ratio change	154	0.26	0.23	0.11	0.30	0.41

Note: The blue dots in the figure represent the out-of-sample annualized Sharpe ratio of the Jensen, Kelly, and Pedersen (2023) 154 portfolios, including the Fama-French-Carhart six factors and a large list of other firm characteristics-based anomaly portfolios. These anomaly portfolios are also organized into 13 thematic categories. The red diamonds represent the Sharpe ratio of the portfolio that optimally combines the original anomaly portfolio with our MPIR strategy. The table at the bottom presents the summary statistics of the changes in the Sharpe ratio between the original anomaly portfolios and the combined portfolios. The expanding windows span from January 2000 to December 2004, and the out-of-sample testing period extends from January 2005 to September 2020.

Table 4. Robustness to skip, lookback period, and normalization



Note: In the top two panels, we report the Sharpe ratio of the MPIR strategy and the average Sharpe ratio change for the 154 anomaly portfolios for different skip and lookback months. Our base specification, which trades against the flows from the past six months, corresponds to the (0,6) position. The top two panels apply normalization by the in-sample reversion ratio, consistent with our base specification, while the bottom two panels do not. The out-of-sample testing period spans from January 2005 to September 2020.

6.4 Robustness

Here, we show that the MPIR strategy is robust to alternative specifications. Our baseline strategy trades against the average factor flows from the past six months. However, there

may be concerns about the accessibility of monthly flow information at the start of the subsequent month. Moreover, the decision to look back six months, instead of a different period, might be questioned. We now demonstrate that our findings remain robust to these concerns.

In the top two panels of Table 4, we report the Sharpe ratio of the MPIR strategy and the average Sharpe ratio change for the 154 anomaly portfolios for different skip and lookback months. For a large set of alternative parameters, the investment results align closely with our baseline specification, which corresponds to the (0,6) position in the table. Importantly, even if we skip two months (i.e., using the flow information up to January at the end of March), our strategy still performs reasonably well.

In the bottom two panels of Table 4, we replicate the analysis from the top two panels without applying normalization based on the in-sample reversion ratio. The investment outcomes are similar, though slightly diminished, indicating that the in-sample normalization is useful.

7 Conclusion

In conclusion, this paper studies the impact of noise trading flows on the cross-section of asset prices in a market where sophisticated investors enforce no-arbitrage. In our model, individual asset flows, aggregated at the factor level, drive fluctuations in factor risk premia, which in turn impact asset prices through beta pricing. This structure fits the reduced-form patterns of how each asset's flow impacts its own price and other assets' prices with only a few factor-level parameters. Furthermore, a trading strategy derived from the model, designed to capitalize on the reversion of factor-level price impacts, delivers strong investment outcomes and enhances the performance of a wide range of anomaly portfolios.

The factor framework we propose is versatile, with the potential to extend beyond the specific factors and flows explored in this study. Importantly, noise trading plays a critical role not only in equity markets but also in treasuries, corporate bonds, and foreign exchange

markets, as evidenced by Vayanos and Vila (2021), Chaudhary, Fu, and Li (2023), and Jeanne and Rose (2002). Even within the equity market, mutual fund flow-induced trading is not the only source of noise trading. For example, Boehmer, Jones, Zhang, and Zhang (2021) identify retail investor trades using publically available U.S. equity transactions data, which could potentially be another source of noise trading. Moreover, while we find good empirical performance using the most canonical Fama-French factors, exploring other risk factors influenced by noise trading could further improve the model fit and investment performance. Modern factor-model techniques, such as those proposed by Kozak, Nagel, and Santosh (2020), offer a promising direction to enhance factor identification within our model.

References

Alekseev, Georgij, Stefano Giglio, Quinn Maingi, Julia Selgrad, and Johannes Stroebel, 2022, A quantity-based approach to constructing climate risk hedge portfolios, Working paper, NYU.

Alvarez, Fernando, and Andrew Atkeson, 2018, The risk of becoming risk averse: A model of asset pricing and trade volumes, Working paper, University of Chicago.

Amihud, Yakov, and Haim Mendelson, 1991, Liquidity, maturity, and the yields on us treasury securities, *Journal of Finance* 46, 1411–1425.

An, Yu, 2023, Flow-based arbitrage pricing theory, Working paper, Johns Hopkins University.

An, Yu, and Zeyu Zheng, 2023, A dynamic factor model of price impacts, Working paper, Johns Hopkins University.

Andrade, Sandro C, Charles Chang, and Mark S Seasholes, 2008, Trading imbalances, predictable reversals, and cross-stock price pressure, *Journal of Financial Economics* 88, 406–423.

Balasubramaniam, Vimal, John Y Campbell, Tarun Ramadorai, and Benjamin Ranish, 2021, Who owns what? A factor model for direct stockholding, *Journal of Finance* Forthcoming.

Barber, Brad M, Xing Huang, Terrance Odean, and Christopher Schwarz, 2022, Attention-induced trading and returns: Evidence from Robinhood users, *Journal of Finance* 77, 3141–3190.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2022a, Ratings-driven demand and systematic price fluctuations, *Review of Financial Studies* 35, 2790–2838.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2022b, What do mutual fund investors really care about? *Review of Financial Studies* 35, 1723–1774.

Boehmer, Ekkehart, Charles M Jones, Xiaoyan Zhang, and Xinran Zhang, 2021, Tracking retail investor activity, *Journal of Finance* 76, 2249–2305.

Bogousslavsky, Vincent, and Dmitriy Muravyev, 2023, Who trades at the close? implications for price discovery and liquidity, *Journal of Financial Markets* 100852.

Boulatov, Alex, Terrence Hendershott, and Dmitry Livdan, 2013, Informed trading and portfolio returns, *Review of Economic Studies* 80, 35–72.

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

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, 2015, Regression discontinuity and the price effects of stock market indexing, *Review of Financial Studies* 28, 212–246.

Chaudhary, Manav, Zhiyu Fu, and Jian Li, 2023, Corporate bond multipliers: Substitutes matter, Working paper, Columbia Business School.

Christoffersen, Susan Kerr, Erfan Danesh, and David K Musto, 2015, Why do institutions delay reporting their shareholdings? Evidence from form 13F, Working paper, University of Toronto.

Cochrane, John H, 2009, *Asset pricing: Revised edition* (Princeton university press).

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

Daniel, Kent D, David Hirshleifer, and Avanidhar Subrahmanyam, 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 56, 921–965.

De Bondt, Werner F.M., and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793–805.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.

Dou, Winston, Leonid Kogan, and Wei Wu, 2022, Common fund flows: Flow hedging and factor pricing, *Journal of Finance* Forthcoming.

Du, Wenxin, Alexander Tepper, and Adrien Verdelhan, 2018, Deviations from covered interest rate parity, *Journal of Finance* 73, 915–957.

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

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

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

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

Gabaix, Xavier, and Ralph SJ Koijen, 2022, In search of the origins of financial fluctuations: The inelastic markets hypothesis, Working paper, Harvard University.

Gibbons, Michael R, Stephen A Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 1121–1152.

Giglio, Stefano, and Dacheng Xiu, 2021, Asset pricing with omitted factors, *Journal of Political Economy* 129, 1947–1990.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of returns and expected returns, *Review of Financial Studies* 27, 714–746.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Hasbrouck, Joel, and Duane J Seppi, 2001, Common factors in prices, order flows, and liquidity, *Journal of Financial Economics* 59, 383–411.

Huang, Shiyang, Yang Song, and Hong Xiang, 2024, Noise trading and asset pricing factors, *Management Science* Forthcoming.

Jeanne, Olivier, and Andrew K Rose, 2002, Noise trading and exchange rate regimes, *Quarterly Journal of Economics* 117, 537–569.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *Journal of Finance* 45, 881–898.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance? *Journal of Finance* 78, 2465–2518.

Kang, Wenjin, K Geert Rouwenhorst, and Ke Tang, 2022, Crowding and factor returns, Working paper, Yale University.

Kelly, Bryan, Semyon Malamud, and Lasse Heje Pedersen, 2023, Principal portfolios, *Journal of Finance* 78, 347–387.

Kelly, Bryan T, Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Kim, Minsoo, 2020, Fund flows, liquidity, and asset prices, Working paper, University of Melbourne.

Koijen, Ralph SJ, and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *Journal of Finance* 73, 1183–1223.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

Lamont, Owen A., and Richard H. Thaler, 2003, Can the market add and subtract? mis-pricing in tech stock carve-outs, *Journal of Political Economy* 111, 227–268.

Lee, Charles MC, Andrei Shleifer, and Richard H Thaler, 1991, Investor sentiment and the closed-end fund puzzle, *Journal of Finance* 46, 75–109.

Li, Jiacui, 2022, What drives the size and value factors? *Review of Asset Pricing Studies* 12, 845–885.

Li, Jiacui, and Zihan Lin, 2022, Prices are less elastic at more aggregate levels, Working paper, University of Utah.

Lo, Andrew W, and Jiang Wang, 2000, Trading volume: definitions, data analysis, and implications of portfolio theory, *Review of Financial Studies* 13, 257–300.

Loewenstein, Mark, and Gregory A. Willard, 2006, The limits of investor behavior, *Journal of Finance* 61, 231–258.

Lou, Dong, 2012, A flow-based explanation for return predictability, *Review of Financial Studies* 25, 3457–3489.

Lou, Dong, Christopher Polk, and Spyros Skouras, 2019, A tug of war: Overnight versus intraday expected returns, *Journal of Financial Economics* 134, 192–213.

Lou, Dong, Christopher Polk, and Spyros Skouras, 2022, The day destroys the night, night extends the days, Working paper, LSE.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *Journal of Finance* 72, 1611–1644.

Nagel, Stefan, 2012, Evaporating liquidity, *Review of Financial Studies* 25, 2005–2039.

Pasquariello, Paolo, and Clara Vega, 2015, Strategic cross-trading in the US stock market, *Review of Finance* 19, 229–282.

Rostek, Marzena J, and Ji Hee Yoon, 2023, Imperfect competition in financial markets: Recent developments, Working paper, University of Wisconsin - Madison.

Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.

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

Train, Kenneth E, 2009, *Discrete choice methods with simulation* (Cambridge university press).

Vayanos, Dimitri, and Jean-Luc Vila, 2021, A preferred-habitat model of the term structure of interest rates, *Econometrica* 89, 77–112.

Warther, Vincent A., 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39, 209–235.

Appendix

The appendices provide additional details and results.

A Technical Details

In this appendix, we provide technical details omitted in the main text.

A.1 Details for Rotation

In this appendix, we present details for rotating factors. The goal is to find a $K \times K$ matrix $\mathbf{O}$, such that the rotated factors defined using

$$(\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \ldots, \tilde{\mathbf{b}}_K) := (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_K)\mathbf{O}, \tag{A.1}$$

$$(\tilde{q}_{1,t}, \tilde{q}_{2,t}, \ldots, \tilde{q}_{K,t})^\top := \mathbf{O}^{-1}(q_{1,t}, q_{2,t}, \ldots, q_{K,t})^\top. \tag{A.2}$$

have uncorrelated fundamental returns and flows, i.e., $\text{cov}(\tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t, \tilde{\mathbf{b}}_j^\top \boldsymbol{\xi}_t) = 0$ and $\text{cov}(\tilde{q}_{k,t}, \tilde{q}_{j,t}) = 0$ for any $k \neq j$. The calculations remain the same whether we use the payoff $\mathbf{X}$ and flow measured in the number of shares, or the fundamental return $\boldsymbol{\xi}_t$ and flow measured in dollar values. For simplicity, we opt for the latter setup as in equation (11).

We write portfolio weights of the K factors as an $N \times K$ matrix $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_K)$. We write the K factors' flows as a $K \times 1$ vector $\mathbf{q}_t = (q_{1,t}, q_{2,t}, \ldots, q_{K,t})^\top$. In matrix form, the conditions for uncorrelated fundamental returns and flows become

$$\mathbf{O}^\top \mathbf{B}^\top \text{var}(\boldsymbol{\xi}_t)\mathbf{B}\mathbf{O} = \mathbf{I}_K, \tag{A.3}$$

$$\mathbf{O}\boldsymbol{\Pi}\mathbf{O}^\top = \text{var}(\mathbf{q}_t), \tag{A.4}$$

where $\boldsymbol{\Pi} = \text{diag}(\pi_1, \pi_2, \ldots, \pi_K)$ is some $K \times K$ diagonal matrix.

To obtain $\mathbf{O}$, we first carry out Cholesky decomposition and obtain

$$\mathbf{B}^\top \mathrm{var}(\boldsymbol{\xi}_t)\mathbf{B} = \mathbf{U}^\top \mathbf{U}, \tag{A.5}$$

where $\mathbf{U}$ is an $K \times K$ upper triangular matrix. Second, we carry out eigenvalue decomposition

$$\left(\mathbf{U}\mathrm{var}(\mathbf{q}_t)\mathbf{U}^\top\right)\mathbf{G} = \mathbf{G}\boldsymbol{\Pi}, \tag{A.6}$$

where $\boldsymbol{\Pi} = \mathrm{diag}(\pi_1, \pi_2, \ldots, \pi_K)$, and $\mathbf{G}$ is an orthogonal $K \times K$ matrix satisfying $\mathbf{G}^{-1} = \mathbf{G}^\top$. We claim that $\mathbf{O} = \mathbf{U}^{-1}\mathbf{G}$ satisfies the conditions (A.3) and (A.4).

Proof. First, we see that

$$\mathbf{O}^\top \mathbf{B}^\top \mathrm{var}(\boldsymbol{\xi}_t)\mathbf{B}\mathbf{O} = \mathbf{G}^\top (\mathbf{U}^\top)^{-1}\mathbf{U}^\top \mathbf{U}\mathbf{U}^{-1}\mathbf{G} = \mathbf{I}_K. \tag{A.7}$$

Second, we have by (A.6),

$$\mathbf{U}\mathrm{var}(\mathbf{q}_t)\mathbf{U}^\top \mathbf{U}\mathbf{O} = \mathbf{U}\mathbf{O}\boldsymbol{\Pi}. \tag{A.8}$$

Eliminating the invertible matrix $\mathbf{U}$ on both sides, we obtain

$$\mathrm{var}(\mathbf{q}_t)\mathbf{U}^\top \mathbf{U}\mathbf{O} = \mathbf{O}\boldsymbol{\Pi}. \tag{A.9}$$

Note that

$$\mathbf{O}^\top \mathbf{U}^\top \mathbf{U}\mathbf{O} = \mathbf{G}^\top \mathbf{G} = \mathbf{I}_K. \tag{A.10}$$

Therefore, we have

$$\mathrm{var}(\mathbf{q}_t)(\mathbf{O}^\top)^{-1} = \mathbf{O}\boldsymbol{\Pi}, \tag{A.11}$$

which proves (A.4). □

A.2 Mean-Variance Optimization for Price Impacts

In this appendix, we present the theory of mean-variance optimization for price impacts. We show that the optimal strategy for capitalizing on flow improves fundamental investing strategies, providing theoretical support for empirical findings in Section 6.3.

We follow the notations in Section 3. Recall that under flow $\mathbf{f}$, the asset returns from time 0 to 1 are denoted as

$$\mathbf{R}(\mathbf{f}) = \left(\frac{X_1}{P_1(\mathbf{f})}, \frac{X_2}{P_2(\mathbf{f})}, \ldots, \frac{X_N}{P_N(\mathbf{f})}\right)^\top. \tag{A.12}$$

In particular, the fundamental returns $\mathbf{R}(\mathbf{0})$ are asset returns when there are no flows.

As a starting point, we introduce the price impact ratio, which quantifies how flows change a portfolio's Sharpe ratio. For any portfolio $\mathbf{c} = (c_1, c_2, \ldots, c_N)^\top$, where c_n is the dollar value invested in asset n when asset prices are $P_n(\mathbf{0})$, we define the *price impact ratio* of portfolio $\mathbf{c}$ in the economy with flow $\mathbf{f}$ as

$$\mathrm{PIR}(\mathbf{c}, \mathbf{f}) := \frac{\mathbf{c}^\top \Delta\mathbf{p}(\mathbf{f})}{\sigma(\mathbf{c}^\top \mathbf{R}(\mathbf{0}))}, \tag{A.13}$$

where $\mathbf{c}^\top \Delta\mathbf{p}(\mathbf{f})$ is the portfolio's price impact and $\sigma(\mathbf{c}^\top \mathbf{R}(\mathbf{0}))$ is the fundamental-return risk. In this definition, the denominator uses $\mathbf{R}(\mathbf{0})$, the assets' fundamental returns when there are no flows, instead of $\mathbf{R}(\mathbf{f})$. This definition ensures that the measurement of portfolio risk is not contaminated by flow-induced changes in time-0 asset prices. The following proposition shows that the price impact ratio is flow-induced changes in the Sharpe ratios.

PROPOSITION A.1. *We have*

$$R_F \mathrm{PIR}(\mathbf{c}, \mathbf{f}) = \mathrm{SR}(\mathbf{c}, \mathbf{0}) - \mathrm{SR}(\mathbf{c}, \mathbf{f}), \tag{A.14}$$

where the Sharpe ratio of portfolio $\mathbf{c}$ *in the economy with flow* $\mathbf{f}$ *is defined in the standard*

way as

$$\mathrm{SR}(\mathbf{c},\mathbf{f}) := \frac{\mathbb{E}[(\mathbf{W}(\mathbf{f})\mathbf{c})^\top(\mathbf{R}(\mathbf{f})-R_F\mathbf{1})]}{\sigma((\mathbf{W}(\mathbf{f})\mathbf{c})^\top\mathbf{R}(\mathbf{f}))}, \tag{A.15}$$

with the $N\times N$ diagonal matrix[31] $\mathbf{W}(\mathbf{f}) := \mathrm{diag}(P_1(\mathbf{f})/P_1(\mathbf{0}), P_2(\mathbf{f})/P_2(\mathbf{0}), \ldots, P_N(\mathbf{f})/P_N(\mathbf{0}))$.

Proof. We note that $\mathbf{W}(\mathbf{f})\mathbf{R}(\mathbf{f}) = \mathbf{R}(\mathbf{0})$. Therefore, we have

$$\begin{aligned}\mathrm{SR}(\mathbf{c},\mathbf{0}) - \mathrm{SR}(\mathbf{c},\mathbf{f}) &= \frac{\mathbb{E}[\mathbf{c}^\top(\mathbf{R}(\mathbf{0})-R_F\mathbf{1})]}{\sigma(\mathbf{c}(\mathbf{f})^\top\mathbf{R}(\mathbf{0}))} - \frac{\mathbb{E}[\mathbf{c}^\top(\mathbf{R}(\mathbf{0})-R_F\mathbf{W}(\mathbf{f})\mathbf{1})]}{\sigma(\mathbf{c}(\mathbf{f})^\top\mathbf{R}(\mathbf{0}))} \\ &= R_F\frac{\mathbb{E}[\mathbf{c}^\top(\mathbf{W}(\mathbf{f})\mathbf{1}-\mathbf{1})]}{\sigma(\mathbf{c}(\mathbf{f})^\top\mathbf{R}(\mathbf{0}))} = R_F\frac{\mathbf{c}^\top\Delta\mathbf{p}(\mathbf{f})}{\sigma(\mathbf{c}^\top\mathbf{R}(\mathbf{0}))} = R_F\mathrm{PIR}(\mathbf{c},\mathbf{f}).\end{aligned} \tag{A.16}$$

□

Next, we consider the mean-variance optimal portfolio of price impacts and show that this portfolio improves fundamental investing strategies. We define the maximum-price-impact-ratio (MPIR) portfolio under flow $\mathbf{f}$ as the portfolio with the maximum amount of price impacts per unit of risks,

$$\tilde{\mathbf{c}}^*(\mathbf{f}) := \mathrm{var}(\mathbf{R}(\mathbf{0}))^{-1}\Delta\mathbf{p}(\mathbf{f}) \in \arg\max_{\mathbf{c}\in\mathbb{R}^N}\mathrm{PIR}(\mathbf{c},\mathbf{f}). \tag{A.17}$$

The maximum-Sharpe-ratio portfolio under flow $\mathbf{f}$ is defined in the standard way as

$$\mathbf{c}^*(\mathbf{f}) := \mathbf{W}(\mathbf{f})^{-1}\mathrm{var}(\mathbf{R}(\mathbf{f}))^{-1}\mathbb{E}[\mathbf{R}(\mathbf{f})-R_F\mathbf{1}] \in \arg\max_{\mathbf{c}\in\mathbb{R}^N}\mathrm{SR}(\mathbf{c},\mathbf{f}). \tag{A.18}$$

The MPIR portfolio $\tilde{\mathbf{c}}^*(\mathbf{f})$ is not simply the portfolio with the largest price impacts but also depends on risks. Intuitively, we want to find two assets that either have opposite price impacts but positively correlated risks (so we can long one and short the other) or have similar price impacts but weakly or negatively correlated risks (so we can diversify). Trading against the MPIR portfolio offers the best price-impact-versus-risk trade-off for a liquidity

[31] Recall that the portfolio weights $\mathbf{c} = (c_1, c_2, \ldots, c_N)^\top$ are in the unit of dollar amounts invested in asset n when asset prices are $P_n(\mathbf{0})$. When asset prices change from $P_n(\mathbf{0})$ to $P_n(\mathbf{f})$ with flow $\mathbf{f}$, the dollar amounts need to change from $\mathbf{c}$ to $\mathbf{W}(\mathbf{f})\mathbf{c}$ for the same portfolio.

provider. Theorem A.1 makes precise this intuition (see Appendix B.1 for a proof).

THEOREM A.1 (**Two-portfolio separation**). *We have*

$$\underbrace{\mathbf{c}^*(\mathbf{f})}_{\substack{\text{max. Sharpe ratio portfolio with flow}}} = \overbrace{\mathbf{c}^*(\mathbf{0})}^{\substack{\text{max. Sharpe ratio portfolio without flow}}} - R_F \underbrace{\tilde{\mathbf{c}}^*(\mathbf{f})}_{\substack{\text{max. price impact ratio portfolio}}}. \tag{A.19}$$

The return volatility of portfolio $\mathbf{c}^*(\mathbf{0})$ *equals the maximum Sharpe ratio without flow. The return volatility of portfolio* $\tilde{\mathbf{c}}^*(\mathbf{f})$ *equals the maximum price impact ratio.*

Equation (A.19) shows that the maximum-Sharpe-ratio portfolio $\mathbf{c}^*(\mathbf{f})$ under flow $\mathbf{f}$ can be separated into two. The first portfolio $\mathbf{c}^*(\mathbf{0})$ maximizes the Sharpe ratio in the same economy but without flow or, equivalently, maximizes the Sharpe ratio driven by the fundamental returns $\mathbf{R}(\mathbf{0})$. The second portfolio $\tilde{\mathbf{c}}^*(\mathbf{f})$ maximizes the price impact ratio under flow $\mathbf{f}$. Intuitively, the fundamental-investing portfolio $\mathbf{c}^*(\mathbf{0})$ is the static mean-variance optimal portfolio that ignores the flow information. Many existing anomaly portfolios fall into this category. On the other hand, the MPIR portfolio $\tilde{\mathbf{c}}^*(\mathbf{f})$, which we empirically construct, optimally times the flow-induced changes in risk premia.

Theorem A.1 shows that longing the fundamental-investing portfolio $\mathbf{c}^*(\mathbf{0})$ and shorting the MPIR portfolio $\tilde{\mathbf{c}}^*(\mathbf{f})$ maximize the overall Sharpe ratio. Empirically, this implies that the Sharpe ratio of an existing anomaly should increase once we combine it with our MPIR strategy. Because of diversification benefits, the two portfolios' risk exposures are roughly proportional to their respective fundamental-based Sharpe ratio and flow-based price impact ratio. That is, in periods with greater noisy flows, marginal investors optimally allocate greater risk exposure to liquidity provision; in periods with smaller flows, marginal investors allocate greater risk exposure to fundamental investing.

A.3 Idiosyncratic Flow and Mean-Variance Optimal Stratgy

In this appendix, we establish the theoretical linkage between the price impacts of idiosyncratic flows and the mean-variance optimal strategy that capitalizes on these flows.

We run the second-stage regression (11), supplementing it with extra terms for idiosyncratic flows,

$$r_{n,t} = \sum_{m=1}^{N} a_{n,m} e_{m,t} + \sum_{k=1}^{K} \lambda_k \tilde{q}_{k,t} \mathrm{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t) + \xi_{n,t}. \tag{A.20}$$

The idiosyncratic flow $e_{m,t}$ of asset m at time t is the residual from the first-stage regression (12). The coefficient $a_{n,m}$ measures the residual price impact on asset n created by the idiosyncratic flow into asset m. Our factor model (11) implies the null hypothesis

$$H_0 : a_{n,m} = 0 \text{ for all } n \text{ and } m. \tag{A.21}$$

That is, the idiosyncratic flow into any asset m should not generate price impacts for any asset n, including itself. Let $\hat{\mathbf{a}}$ be the $N^2 \times 1$ vector of parameter estimates $\hat{a}_{n,m}$. As period T tends to infinity, the asymptotic χ^2 test statistic for the null hypothesis is

$$T \hat{\mathbf{a}}^\top \frac{\mathrm{var}(\hat{\mathbf{a}})^{-1}}{T} \hat{\mathbf{a}} \sim \chi^2_{N^2}. \tag{A.22}$$

To understand the economics of the χ^2 test statistic, we study the mean-variance optimal strategy that capitalizes on flows, as defined in Appendix A.2. Under our factor model (11), the price impact of asset n at time t is $\delta_{n,t} := \sum_{k=1}^{K} \hat{\lambda}_k \tilde{q}_{k,t} \mathrm{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t)$, where we use $\hat{\lambda}_k$ to denote the estimates of λ_k. As defined in (A.13), the price impact ratio of any portfolio $\mathbf{c} \in \mathbb{R}^N$ at time t is $\mathbf{c}^\top \boldsymbol{\delta}_t / \sigma(\mathbf{c}^\top \boldsymbol{\xi}_t)$, where the cross-section of price impact is denoted as $\boldsymbol{\delta}_t = (\delta_{1,t}, \delta_{2,t}, \ldots, \delta_{N,t})^\top$, and the N assets' fundamental returns at time t is denoted as $\boldsymbol{\xi}_t$. The maximum price impact ratio (MPIR) across all portfolios is

$$\max_{\mathbf{c} \in \mathbb{R}^N} \frac{\mathbf{c}^\top \boldsymbol{\delta}_t}{\sigma(\mathbf{c}^\top \boldsymbol{\xi}_t)} = \sqrt{\boldsymbol{\delta}_t^\top \mathrm{var}(\boldsymbol{\xi}_t)^{-1} \boldsymbol{\delta}_t}. \tag{A.23}$$

We define the time-series average of the model-implied squared MPIR as

$$\theta^2 := \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\delta}_t^\top \mathrm{var}(\boldsymbol{\xi}_t)^{-1}\boldsymbol{\delta}_t. \tag{A.24}$$

Similarly, the price impact under the model (A.20) with residual price impacts[32] is $\check{\delta}_{n,t} := \sum_{m=1}^{N} \hat{a}_{n,m}e_{m,t} + \sum_{k=1}^{K} \hat{\lambda}_k \tilde{q}_{k,t}\mathrm{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t)$, with $\check{\boldsymbol{\delta}}_t = (\check{\delta}_{1,t}, \check{\delta}_{2,t}, \ldots, \check{\delta}_{N,t})^\top$. We define the time-series average of the realized squared MPIR as

$$\check{\theta}^2 := \frac{1}{T}\sum_{t=1}^{T} \check{\boldsymbol{\delta}}_t^\top \mathrm{var}(\boldsymbol{\xi}_t)^{-1}\check{\boldsymbol{\delta}}_t. \tag{A.25}$$

The following theorem connects the MPIR to the residual impacts of idiosyncratic flows (see Appendix B.3 for a proof).

THEOREM A.2. *Assuming that the regressors in* (A.20) *are observed without standard errors and the fundamental return* $\boldsymbol{\xi}_t$ *is i.i.d. over time, we have, almost surely as* T *tends to infinity,*

$$\hat{\mathbf{a}}^\top \frac{\mathrm{var}(\hat{\mathbf{a}})^{-1}}{T}\hat{\mathbf{a}} = \check{\theta}^2 - \theta^2. \tag{A.26}$$

Equation (A.26) shows that as period T tends to infinity, the χ^2 test statistic for residual price impacts in (A.22) equals the difference between the realized and model-implied average squared MPIR. This result generalizes Gibbons, Ross, and Shanken (1989), who show that the χ^2 test statistic for anomaly expected returns of idiosyncratic risks equals the difference between the realized and model-implied squared maximum Sharpe ratio. Simply put, Gibbons, Ross, and Shanken (1989) connect the idiosyncratic risks to the maximum Sharpe ratio, whereas we connect the idiosyncratic flows to the maximum price impact ratio. We assume observed regressors and i.i.d. fundamental returns to avoid econometric complications. We leave the generalization of the Shanken (1992) correction for future research.

[32] Appendix B.2 proves that the estimated $\hat{\lambda}_k$ remain unchanged in regressions (11) and (A.20).

A.4 Details for the MPIR strategy

In this appendix, we provide details on constructing the MPIR strategy and how to combine it with existing anomaly portfolios.

First, using equation (18), we define the staggered MPIR strategy in month s as

$$\tilde{\mathbf{c}}_s^{\{t\}} := \sum_{k\in\{\text{MKT,SMB,HML}\}} -\lambda_k^{\{t\}} \bar{q}_{k,s}^{\{t\}} \tilde{\mathbf{b}}_k^{\{t\}}. \tag{A.27}$$

The superscript $\{t\}$ indicates that $\tilde{\mathbf{c}}_s^{\{t\}}$ is estimated using the training window up to month t. The term $\lambda_k^{\{t\}}$ represents the estimated price of flow-induced risk of factor k, $\bar{q}_{k,s}^{\{t\}}$ represents the average flow into factor k over past six months $s-5, s-4, \ldots, s$, and $\tilde{\mathbf{b}}_k^{\{t\}}$ represents the portfolio weights of factor k.

Second, we compute the model-implied price impact ratio and actual price reversion of the staggered MPIR portfolio $\tilde{\mathbf{c}}_s^{\{t\}}$. By equation (A.13), the price impact ratio in month s is

$$\kappa_s^{\{t\}} := \sqrt{\sum_{k\in\{\text{MKT,SMB,HML}\}} \left(\lambda_k^{\{t\}} \bar{q}_{k,s}^{\{t\}}\right)^2}, \tag{A.28}$$

where we have applied the condition $(\tilde{\mathbf{b}}_k^{\{t\}})^\top \text{var}(\boldsymbol{\xi})^{\{t\}} \tilde{\mathbf{b}}_k^{\{t\}} = 1$ and $(\tilde{\mathbf{b}}_k^{\{t\}})^\top \text{var}(\boldsymbol{\xi})^{\{t\}} \tilde{\mathbf{b}}_l^{\{t\}} = 0$ for $k \neq l$ to simplify the expression. Similarly, the price reversion of the staggered MPIR portfolio $\tilde{\mathbf{c}}_s^{\{t\}}$ in month $s+1$, normalized by its fundamental risk, is

$$\check{\kappa}_s^{\{t\}} := \frac{(\tilde{\mathbf{c}}_s^{\{t\}})^\top \tilde{\mathbf{r}}_{s+1}}{\sqrt{(\tilde{\mathbf{c}}_s^{\{t\}})^\top \text{var}(\boldsymbol{\xi})^{\{t\}} \tilde{\mathbf{c}}_s^{\{t\}}}} = \frac{(\tilde{\mathbf{c}}_s^{\{t\}})^\top \tilde{\mathbf{r}}_{s+1}}{\sqrt{\sum_{k\in\{\text{MKT,SMB,HML}\}} \left(\lambda_k^{\{t\}} \bar{q}_{k,s}^{\{t\}}\right)^2}}, \tag{A.29}$$

where $\tilde{\mathbf{r}}_{s+1}$ is the demeaned return of the 25 test assets in month $s+1$.

Third, we define the ratio of the average reversion to the average model-implied price impact as

$$\text{REV}^{\{t\}} = \frac{\sum_{s=6}^{t-1} \check{\kappa}_s^{\{t\}}}{\sum_{s=6}^{t-1} \kappa_s^{\{t\}}}. \tag{A.30}$$

The summation over month s starts from 6 because $\bar{q}_{k,s}^{\{t\}}$ is the average flow over the past six months and ends at $t-1$ because $\mathbf{r}_t$ is the last observable return using the training window ending in month t.

Fourth, as discussed in the main text, we normalize the MPIR strategy using the in-sample estimated reversion ratio, which is confined between 0 and 1. The formula is:

$$\tilde{\mathbf{c}}_t^* = \max\left(\min\left(\mathrm{REV}^{\{t\}}, 1\right), 0\right)\tilde{\mathbf{c}}_t^{\{t\}}. \tag{A.31}$$

The MPIR strategy's excess return in month $t+1$ is defined as

$$\tilde{r}_{t+1}^* = (\tilde{\mathbf{c}}_t^*)^\top(\mathbf{r}_{t+1} - r_{F,t+1}), \tag{A.32}$$

where $\mathbf{r}_{t+1}$ is the 25 test assets' return in month $t+1$ and $r_{F,t+1}$ is the net risk-free rate.

Finally, we combine the MPIR strategy with each of the anomaly portfolios using equation (19). The mixing ratio is given by

$$w_{j,t} := \max\left(\frac{\mathrm{VOL}_j^{\{t\}}}{\mathrm{SR}_j^{\{t\}}}, 0\right)(1 + r_{F,t+1}), \tag{A.33}$$

according to Theorem A.1. The term $\mathrm{SR}_j^{\{t\}}$ represents the Sharpe ratio of anomaly portfolio j, and $\mathrm{VOL}_j^{\{t\}}$ represents the return standard deviation. Both terms are calculated utilizing the same training windows as those used in the MPIR estimation. We normalize the MPIR strategy by $\mathrm{VOL}_j^{\{t\}}/\mathrm{SR}_j^{\{t\}}$, rather than normalizing portfolio j by $\mathrm{SR}_j^{\{t\}}/\mathrm{VOL}_j^{\{t\}}$. We do so to avoid normalizing portfolio j by the inverse of volatility in the time series, because Moreira and Muir (2017) show that such normalization per se increases the Sharpe ratio. Because the estimated Sharpe ratio $\mathrm{SR}_j^{\{t\}}$ could be negative, we bound the scaling by zero. That is, we leave $r_{j,t+1}$ unchanged if portfolio j has a negative historical Sharpe ratio up to month t.

B Proofs

In this appendix, we provide proof.

B.1 Proof of Theorem A.1

We note that $\mathbf{W}(\mathbf{f})\mathbf{R}(\mathbf{f}) = \mathbf{R}(\mathbf{0})$. Therefore, we have

$$\text{var}(\mathbf{R}(\mathbf{f}))^{-1} = \mathbf{W}(\mathbf{f})\text{var}(\mathbf{R}(\mathbf{0}))^{-1}\mathbf{W}(\mathbf{f}). \tag{A.34}$$

Second, we have

$$\mathbf{R}(\mathbf{0}) - R_F\mathbf{1} + R_F\mathbf{1} = \mathbf{W}(\mathbf{f})(\mathbf{R}(\mathbf{f}) - R_F\mathbf{1} + R_F\mathbf{1}), \tag{A.35}$$

which simplifies to

$$\mathbf{R}(\mathbf{f}) - R_F\mathbf{1} = \mathbf{W}(\mathbf{f})^{-1}(\mathbf{R}(\mathbf{0}) - R_F\mathbf{1} - R_F\Delta\mathbf{p}(\mathbf{f})). \tag{A.36}$$

Taking expectations on both sides, we have

$$\mathbb{E}[\mathbf{R}(\mathbf{f})] - R_F\mathbf{1} = \mathbf{W}(\mathbf{f})^{-1}(\mathbb{E}[\mathbf{R}(\mathbf{0})] - R_F\mathbf{1} - R_F\Delta\mathbf{p}(\mathbf{f})). \tag{A.37}$$

Therefore, by equations (A.34) and (A.37), we have

$$\begin{aligned}\mathbf{c}^*(\mathbf{f}) =& \mathbf{W}(\mathbf{f})^{-1}\text{var}(\mathbf{R}(\mathbf{f}))^{-1}(\mathbb{E}[\mathbf{R}(\mathbf{f})] - R_F\mathbf{1}) \\ =& \text{var}(\mathbf{R}(\mathbf{0}))^{-1}(\mathbb{E}[\mathbf{R}(\mathbf{0})] - R_F\mathbf{1} - R_F\Delta\mathbf{p}(\mathbf{f})) = \mathbf{c}^*(\mathbf{0}) - R_F\tilde{\mathbf{c}}^*(\mathbf{f}).\end{aligned} \tag{A.38}$$

The return volatility of portfolio $\mathbf{c}^*(\mathbf{0})$ is

$$\sigma(\mathbf{c}^*(\mathbf{0})^\top \mathbf{R}(\mathbf{0})) = \sqrt{\mathbf{c}^*(\mathbf{0})^\top \text{var}(\mathbf{R}(\mathbf{0}))\mathbf{c}^*(\mathbf{0})} = \sqrt{\mathbb{E}[\mathbf{R}(\mathbf{0}) - R_F\mathbf{1}]^\top \text{var}(\mathbf{R}(\mathbf{0}))^{-1}\mathbb{E}[\mathbf{R}(\mathbf{0}) - R_F\mathbf{1}]}, \tag{A.39}$$

which equals the maximum Sharpe ratio without flow by definition (A.18). Similarly, the return volatility of portfolio $\tilde{\mathbf{c}}^*(\mathbf{f})$ is

$$\sigma(\tilde{\mathbf{c}}^*(\mathbf{f})^\top \mathbf{R}(\mathbf{0})) = \sqrt{\tilde{\mathbf{c}}^*(\mathbf{f})^\top \text{var}(\mathbf{R}(\mathbf{0}))\tilde{\mathbf{c}}^*(\mathbf{f})} = \sqrt{\Delta\mathbf{p}(\mathbf{f})^\top \text{var}(\mathbf{R}(\mathbf{0}))^{-1}\Delta\mathbf{p}(\mathbf{f})}, \tag{A.40}$$

which equals the maximum price impact ratio by definition (A.17).

B.2 Proof of the Simplifying Regression

We show that the panel regression (A.20),

$$r_{n,t} = \sum_{m=1}^{N} a_{n,m}e_{m,t} + \sum_{k=1}^{K} \lambda_k \tilde{q}_{k,t}\text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t) + \xi_{n,t}, \tag{A.41}$$

can be separated into two regressions. The first asset-by-asset time-series regression

$$r_{n,t} = \sum_{m=1}^{N} a_{n,m}e_{m,t} + \xi_{n,t}. \tag{A.42}$$

obtains the same $a_{n,m}$ as regression (A.41). The second panel regression

$$r_{n,t} = \sum_{k=1}^{K} \lambda_k \tilde{q}_{k,t}\text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t) + \xi_{n,t}, \tag{A.43}$$

obtains the same λ_k as regression (A.41).

To see this fact, note that because the idiosyncratic flow $e_{m,t}$ is the residual of the first-stage regression (12), we know by construction that $\sum_{t=1}^{T} q_{k,t}e_{m,t} = 0$. Because each $\tilde{q}_{k,t}$ is a linear combination of $q_{1,t}, q_{2,t}, \ldots, q_{K,t}$, we know that $\sum_{t=1}^{T} \tilde{q}_{k,t}e_{m,t} = 0$. We rewrite the panel

regression (A.41) in vector form as

$$\mathbf{r} = \sum_{n=1}^{N}\sum_{m=1}^{N} a_{n,m}\mathbf{e}_{n,m} + \sum_{k=1}^{K}\lambda_k \mathbf{y}_k + \boldsymbol{\xi}, \tag{A.44}$$

where the $NT \times 1$ vector $\mathbf{r}$ is

$$\mathbf{r} = (r_{1,1}, r_{1,2}, \ldots, r_{1,T}, r_{2,1}, r_{2,2}, \ldots, r_{2,T}, \ldots, r_{N,1}, r_{N,2}, \ldots, r_{N,T})^\top. \tag{A.45}$$

Each vector $\mathbf{e}_{n,m}$ is an $NT \times 1$ vector with only the $(n-1)T+1$-th to nT-th entry ranging from $e_{m,1}$ to $e_{m,T}$ and all other entries equaling zero. Each $\mathbf{y}_k$ is an $NT \times 1$ vector

$$\mathbf{y}_k = \left(\underbrace{\tilde{q}_{k,1}\mathrm{cov}(\xi_{1,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t), \ldots, \tilde{q}_{k,T}\mathrm{cov}(\xi_{1,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t)}_{T \text{ terms}}, \ldots, \underbrace{\tilde{q}_{k,1}\mathrm{cov}(\xi_{N,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t), \ldots, \tilde{q}_{k,T}\mathrm{cov}(\xi_{N,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t)}_{T \text{ terms}} \right)^\top. \tag{A.46}$$

The vector $\boldsymbol{\xi}$ simply stacks all error terms $\xi_{n,t}$.

Note that, for any n, m, and k, we have

$$\mathbf{e}_{n,m}^\top \mathbf{y}_k = \mathrm{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t) \sum_{t=1}^{T} \tilde{q}_{k,t} e_{m,t} = 0, \tag{A.47}$$

where the last equality uses the first step in the proof. As a result, to estimate coefficient $a_{n,m}$ and λ_k, it suffices to run separate regressions

$$\mathbf{r} = \sum_{n=1}^{N}\sum_{m=1}^{N} a_{n,m}\mathbf{e}_{n,m} + \boldsymbol{\xi} \text{ and } \mathbf{r} = \sum_{k=1}^{K}\lambda_k \mathbf{y}_k + \boldsymbol{\xi}. \tag{A.48}$$

This first regression further reduces to the asset-by-asset time-series regression (A.42).

B.3 Proof of Theorem A.2

First, we simplify the χ^2 test statistics in (A.26). We write a $1 \times N$ vector $\mathbf{e}_t = (e_{1,t}, e_{2,t}, \ldots, e_{N,t})$ and an $N \times 1$ vector $\mathbf{a}_n = (a_{n,1}, a_{n,2}, \ldots, a_{n,N})^\top$, and we write regression (A.20) as

$$r_{n,t} = \mathbf{e}_t \mathbf{a}_n + \sum_{k=1}^{K} \lambda_k \tilde{q}_{k,t} \text{cov}(\xi_{n,t}, \tilde{\mathbf{b}}_k^\top \boldsymbol{\xi}_t) + \xi_{n,t}. \tag{A.49}$$

We define the $T \times N$ matrix $\mathbf{e} = (\mathbf{e}_1; \mathbf{e}_2; \ldots; \mathbf{e}_T)$ and the $N \times 1$ vector $\boldsymbol{\xi}_n = (\xi_{n,1}, \xi_{n,2}, \ldots, \xi_{n,T})^\top$. As shown in Appendix B.2, we can run an asset-by-asset time-series regression to obtain the point estimator of $\mathbf{a}_n$ as

$$\hat{\mathbf{a}}_n = (\mathbf{e}^\top \mathbf{e})^{-1} \mathbf{e}^\top \begin{pmatrix} r_{n,1} \\ r_{n,2} \\ \ldots \\ r_{n,T} \end{pmatrix} = \mathbf{a}_n + (\mathbf{e}^\top \mathbf{e})^{-1} \mathbf{e}^\top \boldsymbol{\xi}_n, \tag{A.50}$$

where we use the fact that, for any $n = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, K$, $\sum_{t=1}^{T} e_{n,t} \tilde{q}_{k,t} = 0$, because $e_{n,t}$ is the residual from the first-stage regression.

Therefore, we have, for any m and n,

$$\text{cov}(\hat{\mathbf{a}}_n, \hat{\mathbf{a}}_m) = (\mathbf{e}^\top \mathbf{e})^{-1} \mathbf{e}^\top \text{cov}(\boldsymbol{\xi}_n, \boldsymbol{\xi}_m) \mathbf{e} (\mathbf{e}^\top \mathbf{e})^{-1} = (\mathbf{e}^\top \mathbf{e})^{-1} \boldsymbol{\Sigma}_\xi(n, m), \tag{A.51}$$

because $\boldsymbol{\xi}_t$ is i.i.d. over time. The term $\boldsymbol{\Sigma}_\xi(n, m)$ is the (n, m)-th element of the cross-sectional variance-covariance matrix of $\boldsymbol{\xi}_t$. When constructing the χ^2 test statistic, we use the asymptotically consistent sample covariance matrix $\hat{\boldsymbol{\Sigma}}_\xi$ for the true $\boldsymbol{\Sigma}_\xi$. We denote $\hat{\mathbf{a}} = (\hat{\mathbf{a}}_1; \hat{\mathbf{a}}_2; \ldots; \hat{\mathbf{a}}_N)$ as the $N^2 \times 1$ vector of parameter estimates. By equation (A.51), we have

$$\text{var}(\hat{\mathbf{a}}) = \hat{\boldsymbol{\Sigma}}_\xi \otimes (\mathbf{e}^\top \mathbf{e})^{-1}, \tag{A.52}$$

where $\otimes$ represents the Kronecker product. Therefore, we have

$$\hat{\mathbf{a}}^\top \frac{\text{var}(\hat{\mathbf{a}})^{-1}}{T} \hat{\mathbf{a}} = \hat{\mathbf{a}}^\top \left(\hat{\mathbf{\Sigma}}_\xi^{-1} \otimes (\mathbf{e}^\top \mathbf{e}/T) \right) \hat{\mathbf{a}}. \tag{A.53}$$

Under the null hypothesis of $\mathbf{a} = \mathbf{0}$, we have

$$\begin{aligned}
\hat{\mathbf{a}}^\top \left(\hat{\mathbf{\Sigma}}_\xi^{-1} \otimes (\mathbf{e}^\top \mathbf{e}/T) \right) \hat{\mathbf{a}} &= \sum_{n=1}^{N} \sum_{m=1}^{N} \left((\mathbf{e}^\top \mathbf{e})^{-1} \mathbf{e}^\top \boldsymbol{\xi}_n \right)^\top \hat{\mathbf{\Sigma}}_\xi^{-1}(n,m) (\mathbf{e}^\top \mathbf{e}/T) (\mathbf{e}^\top \mathbf{e})^{-1} \mathbf{e}^\top \boldsymbol{\xi}_m & \text{(A.54)} \\
&= \frac{1}{T} \sum_{n=1}^{N} \sum_{m=1}^{N} \hat{\mathbf{\Sigma}}_\xi^{-1}(n,m) \boldsymbol{\xi}_n^\top \mathbf{e} (\mathbf{e}^\top \mathbf{e})^{-1} \mathbf{e}^\top \boldsymbol{\xi}_m \\
&= \frac{1}{T} \sum_{n=1}^{N} \sum_{m=1}^{N} \hat{\mathbf{\Sigma}}_\xi^{-1}(n,m) \boldsymbol{\psi}_n^\top \boldsymbol{\psi}_m & \text{(A.55)} \\
&= \frac{1}{T} \begin{pmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 \\ \cdots \\ \boldsymbol{\psi}_N \end{pmatrix}^\top \left(\hat{\mathbf{\Sigma}}_\xi^{-1} \otimes \mathbf{I}_T \right) \begin{pmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 \\ \cdots \\ \boldsymbol{\psi}_N \end{pmatrix}, & \text{(A.56)}
\end{aligned}$$

where we define

$$\boldsymbol{\psi}_n = \mathbf{e} (\mathbf{e}^\top \mathbf{e})^{-1} \mathbf{e}^\top \boldsymbol{\xi}_n, \tag{A.57}$$

as the projection of $\boldsymbol{\xi}_n$ onto the idiosyncratic flow space. In step (A.54), we use block matrix multiplication for every N elements and $\hat{\mathbf{\Sigma}}_\xi^{-1}(n,m)$ is the (n,m)-th element of $\hat{\mathbf{\Sigma}}_\xi^{-1}$. Step (A.55) is because the projection matrix $\mathbf{e}(\mathbf{e}^\top \mathbf{e})^{-1} \mathbf{e}^\top$ is idempotent. In step (A.56), we use the block-matrix multiplication in the reverse direction, with each $\boldsymbol{\psi}_n$ as a $T \times 1$ vector.

We define $\boldsymbol{\psi}_t = (\psi_{1,t}, \psi_{2,t}, \ldots, \psi_{N,t})^\top$. In this way, $\boldsymbol{\psi}_n$ is the time-series variation in $\psi_{n,t}$ for a given asset n, and $\boldsymbol{\psi}_t$ is the cross-sectional variation in $\psi_{n,t}$ for a given time t. By

rearranging $\boldsymbol{\psi}_n$ into $\boldsymbol{\psi}_t$, we have

$$\frac{1}{T}\begin{pmatrix}\boldsymbol{\psi}_1\\ \boldsymbol{\psi}_2\\ \cdots\\ \boldsymbol{\psi}_N\end{pmatrix}^{\top}\left(\hat{\boldsymbol{\Sigma}}_{\xi}^{-1}\otimes\mathbf{I}_T\right)\begin{pmatrix}\boldsymbol{\psi}_1\\ \boldsymbol{\psi}_2\\ \cdots\\ \boldsymbol{\psi}_N\end{pmatrix}=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t^{\top}\hat{\boldsymbol{\Sigma}}_{\xi}^{-1}\boldsymbol{\psi}_t. \tag{A.58}$$

Because $\boldsymbol{\xi}_t$ is i.i.d. over time, the strong law of large numbers implies that the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{\xi}$ converges to the true $\boldsymbol{\Sigma}_{\xi}$ almost surely as T tends to infinity. Therefore, in the limit of T tending to infinity, we have almost surely,

$$\hat{\mathbf{a}}^{\top}\frac{\mathrm{var}(\hat{\mathbf{a}})^{-1}}{T}\hat{\mathbf{a}}=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t^{\top}\mathrm{var}(\boldsymbol{\xi}_t)^{-1}\boldsymbol{\psi}_t. \tag{A.59}$$

Next, we transform the squared MPIR in (A.26). We have defined $\check{\boldsymbol{\delta}}_t = (\check{\delta}_{1,t},\check{\delta}_{2,t},\ldots,\check{\delta}_{N,t})^{\top}$ as the cross-section of price impacts at time t. We now define $\check{\boldsymbol{\delta}}_n = (\check{\delta}_{n,1},\check{\delta}_{n,2},\ldots,\check{\delta}_{n,T})^{\top}$. Using equations (A.50) and (A.57), we have

$$\check{\boldsymbol{\delta}}_n=\mathbf{e}(\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\boldsymbol{\xi}_n+\sum_{k=1}^{K}\hat{\lambda}_k\mathrm{cov}(\xi_{n,t},\tilde{\mathbf{b}}_k^{\top}\boldsymbol{\xi}_t)\tilde{\mathbf{q}}_k=\boldsymbol{\psi}_n+\sum_{k=1}^{K}\hat{\lambda}_k\mathrm{cov}(\xi_{n,t},\tilde{\mathbf{b}}_k^{\top}\boldsymbol{\xi}_t)\tilde{\mathbf{q}}_k, \tag{A.60}$$

where $\tilde{\mathbf{q}}_k=(\tilde{q}_{k,1},\tilde{q}_{k,2},\ldots,\tilde{q}_{k,T})^{\top}$. In this way, $\check{\boldsymbol{\delta}}_n$ is the time-series variation in $\check{\delta}_{n,t}$ for a given asset n, and $\check{\boldsymbol{\delta}}_t$ is the cross-sectional variation in $\check{\delta}_{n,t}$ for a given time t. Thus, we have

$$\check{\boldsymbol{\delta}}_t=\boldsymbol{\psi}_t+\mathrm{var}(\boldsymbol{\xi}_t)\sum_{k=1}^{K}\hat{\lambda}_k\tilde{q}_{k,t}\tilde{\mathbf{b}}_k. \tag{A.61}$$

The realized squared MPIR at time t is

$$\begin{aligned}
&\check{\boldsymbol{\delta}}_t^\top \mathrm{var}(\boldsymbol{\xi}_t)^{-1}\check{\boldsymbol{\delta}}_t \\
=&\left(\boldsymbol{\psi}_t^\top + \sum_{k=1}^K \hat{\lambda}_k \tilde{q}_{k,t}\tilde{\mathbf{b}}_k^\top \mathrm{var}(\boldsymbol{\xi}_t)\right)\mathrm{var}(\boldsymbol{\xi}_t)^{-1}\left(\boldsymbol{\psi}_t + \mathrm{var}(\boldsymbol{\xi}_t)\sum_{k=1}^K \hat{\lambda}_k \tilde{q}_{k,t}\tilde{\mathbf{b}}_k\right) \\
=&\boldsymbol{\psi}_t^\top \mathrm{var}(\boldsymbol{\xi}_t)^{-1}\boldsymbol{\psi}_t + 2\boldsymbol{\psi}_t^\top \sum_{k=1}^K \hat{\lambda}_k \tilde{q}_{k,t}\tilde{\mathbf{b}}_k + \left(\sum_{k=1}^K \hat{\lambda}_k \tilde{q}_{k,t}\tilde{\mathbf{b}}_k^\top\right)\mathrm{var}(\boldsymbol{\xi}_t)\sum_{k=1}^K \hat{\lambda}_k \tilde{q}_{k,t}\tilde{\mathbf{b}}_k \\
=&\boldsymbol{\psi}_t^\top \mathrm{var}(\boldsymbol{\xi}_t)^{-1}\boldsymbol{\psi}_t + 2\boldsymbol{\psi}_t^\top \sum_{k=1}^K \hat{\lambda}_k \tilde{q}_{k,t}\tilde{\mathbf{b}}_k + \sum_{k=1}^K \hat{\lambda}_k^2 \tilde{q}_{k,t}^2, \qquad \text{(A.62)}
\end{aligned}$$

where, in the last step, we use $\tilde{\mathbf{B}}^\top \mathrm{var}(\boldsymbol{\xi}_t)\tilde{\mathbf{B}} = \mathbf{I}_K$. The time-series average is

$$\begin{aligned}
\check{\theta}^2 =& \frac{1}{T}\sum_{t=1}^T \check{\boldsymbol{\delta}}_n^\top \mathrm{var}(\boldsymbol{\xi}_t)^{-1}\check{\boldsymbol{\delta}}_n \\
=& \frac{1}{T}\sum_{t=1}^T \boldsymbol{\psi}_t^\top \mathrm{var}(\boldsymbol{\xi}_t)^{-1}\boldsymbol{\psi}_t + \frac{2}{T}\sum_{t=1}^T \boldsymbol{\psi}_t^\top \sum_{k=1}^K \hat{\lambda}_k \tilde{q}_{k,t}\tilde{\mathbf{b}}_k + \frac{1}{T}\sum_{t=1}^T\sum_{k=1}^K \hat{\lambda}_k^2 \tilde{q}_{k,t}^2. \qquad \text{(A.63)}
\end{aligned}$$

Note that for any $n = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, K$, we have

$$\begin{aligned}
\sum_{t=1}^T \tilde{q}_{k,t}\psi_{n,t} =& \sum_{t=1}^T \tilde{q}_{k,t}\mathbf{e}_t(\mathbf{e}^\top\mathbf{e})^{-1}\mathbf{e}^\top\boldsymbol{\xi}_n \\
=& \left(\sum_{t=1}^T \tilde{q}_{k,t}e_{1,t}, \sum_{t=1}^T \tilde{q}_{k,t}e_{2,t}, \ldots, \sum_{t=1}^T \tilde{q}_{k,t}e_{N,t}\right)(\mathbf{e}^\top\mathbf{e})^{-1}\mathbf{e}^\top\boldsymbol{\xi}_n = 0. \qquad \text{(A.64)}
\end{aligned}$$

Therefore, we know that

$$\check{\theta}^2 = \frac{1}{T}\sum_{t=1}^T \boldsymbol{\psi}_t^\top \mathrm{var}(\boldsymbol{\xi}_t)^{-1}\boldsymbol{\psi}_t + \frac{1}{T}\sum_{t=1}^T\sum_{k=1}^K \hat{\lambda}_k^2 \tilde{q}_{k,t}^2. \qquad \text{(A.65)}$$

A similar calculation gives

$$\theta^2 = \frac{1}{T}\sum_{t=1}^T\sum_{k=1}^K \hat{\lambda}_k^2 \tilde{q}_{k,t}^2. \qquad \text{(A.66)}$$

Therefore, we have

$$\check{\theta}^2 - \theta^2 = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t^\top \text{var}(\boldsymbol{\xi}_t)^{-1}\boldsymbol{\psi}_t. \tag{A.67}$$

Using (A.59), we have almost surely in the limit of T tending to infinity,

$$\hat{\mathbf{a}}^\top \frac{\text{var}(\hat{\mathbf{a}})^{-1}}{T}\hat{\mathbf{a}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\psi}_t^\top \text{var}(\boldsymbol{\xi}_t)^{-1}\boldsymbol{\psi}_t = \check{\theta}^2 - \theta^2. \tag{A.68}$$

C Construction and Cleaning of Mutual Fund Flows

In this appendix, we present details involved in constructing and cleaning mutual fund flows.

Our primary data source is the CRSP Survivorship-Bias-Free Mutual Fund database. We start with all funds' return and total net assets (TNA) data at the share-class level. A mutual fund may include multiple share classes. We first drop observations with no valid CRSP identifier, `crsp_fundno`. A few fund-share classes report multiple TNAs in a given month. These are likely data duplicates, so we keep only the first observation of the month. We end up with 8,591,018 share-class×month observations. In what follows, we discuss the cleaning steps for returns and TNA at the share-class level. After cleaning, we aggregate the share-class level data to the fund level.

C.1 Return Cleaning

We first correct data errors in monthly net returns, `mret`.

First, we address extremely positive returns. We study the case in which a particular fund share has returns greater than 100% and has existed for more than one year.[33] We manually check the entire time series of each share class in this subsample. Most of these extreme returns reflect misplaced decimal points, which confound returns in decimal and percentage formats. For these cases, we divide the faulty returns by 100.

[33] We use the one-year threshold because mutual fund return and TNA during the first year are sometimes inaccurate in the CRSP database. For example, return and TNA can be stale or reported using a placeholder number such as 0.1. We address these issues by cross-checking with the alternative database.

Second, we address extreme negative returns. Similarly, we study the case in which a particular fund share has existed for more than one year and has returns lower than -50%. With extremely negative returns, we need to distinguish data errors from significantly negative returns before a fund's closure. Thus, we manually check only the subsample of negative returns that occur at least one year prior to the last observation of a closed fund. We manually check whether these extreme returns reflect data-input errors for this subsample. For the cases with misplaced decimal points, we divide the faulty returns by 100.

C.2 TNA Cleaning

Unlike many prior studies that construct percentage mutual fund flows, we study dollar-value flows to preserve the cross-sectional relative magnitudes. The mutual fund size distribution features a very long right tail. Winsorizing the extreme dollar-value TNA likely removes both valid large values and input errors, generating significant bias. We devise an algorithm to identify and correct erroneous observations of TNA:

1. Using the sample with corrected returns, we calculate dollar flows as in (13) at the share-class level.

2. We study the top and bottom 0.5% of all dollar flows.[34] We manually check this subsample's TNA time series of all share classes. We identify several common errors:

    - Misplaced decimal points (usually by hundredths or thousandths).
    - Stale TNA observations from CRSP, typically when a fund reorganizes its share class offering (e.g., adding a new share class and moving assets into a single share class).
    - CRSP sometimes sets TNA $= 0.1$ for the first few months of a new fund or a new share class.

[34]The choice of the top and bottom 0.5% is motivated by the distribution of dollar flows, where most extreme values tend to occur at these tails.

We correct the misplaced decimal issue. For funds suffering from the latter two problems, we obtain their TNA from Morningstar.[35] Morningstar's TNA data (`Net_Assets_ShareClass_Monthly`) suffer to a lesser extent from these issues than CRSP's TNA data. We conclude by further cross-checking other third-party vendors (e.g., Yahoo Finance and Bloomberg Terminal). Hence, whenever a fund's CRSP TNA deviates more than 50% from its Morningstar TNA, we use the Morningstar TNA.

3. We repeat the previous steps one more time to ensure that we have accounted for most, if not all, extreme errors.

4. We compare our cleaned TNA with total assets (`assets`) from Thomson/Refinitiv Holdings data.[36] Following Coval and Stafford (2007) and Lou (2012), we drop observations whenever our cleaned TNA deviate more than 50% from `assets` from Thomson/Refinitiv.

Using cleaned return and TNA data, we calculate dollar flows at the share-class level using equation (13). We obtain a fund's flows by adding up the flows of all share classes in the same fund. The final sample contains 1,613,579 fund×month observations.

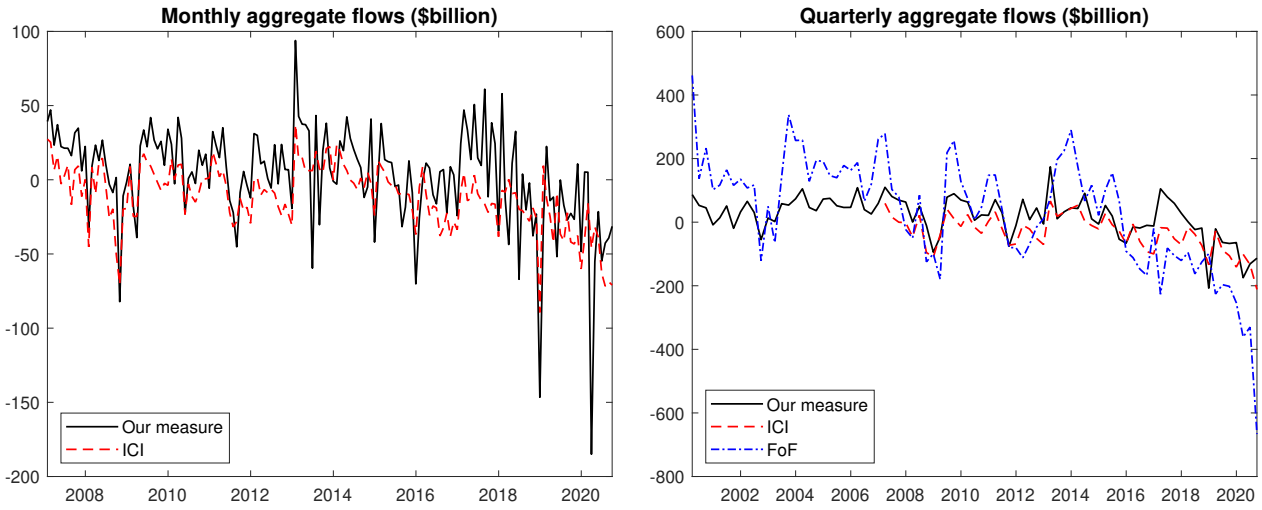
C.3 Cross-Validating the Data-Cleaning Procedure

We cross-validate our data-cleaning procedure. We compute the aggregate mutual fund flows in dollar amounts each month. We compare our aggregate flow measures with alternative sources, including the Investment Company Institute (ICI) and the Flow of Funds (FoF).

The ICI provides aggregate monthly mutual fund flows. We obtain a version of ICI aggregate flows data from 2007 to 2020. We use the ICI's Total Equity mutual fund flows, which feature a close coverage scope to mutual funds in our sample. The FoF data (now known as "Financial Accounts of the United States - Z.1") are published quarterly by the

[35] We merge the CRSP and Morningstar databases using a fund's ticker.
[36] We merge the two databases via the linking table MFLINKS, which WRDS provides.

Figure A.1. Time series of aggregate mutual fund flows from various sources



Note: The left panel plots the monthly time series of our measure of aggregate mutual fund flows and Investment Company Institute (ICI) flows. The right panel plots the quarterly time series of our measure, ICI flows, and Flow of Funds (FoF) flows.

Federal Reserve Board. We use mutual fund flow (Line 28) of Corporate Equities (Table 223) and unadjusted flows (FU). We use the December 2021 vintage of the data because the Federal Reserve revises historical FoF data every quarter.

Figure A.1 plots the time series of aggregate mutual fund flows from various sources. The left panel shows the monthly time series of our measure and ICI flow, and the right panel shows the quarterly time series of all three sources. Our measure of aggregate mutual fund flows is broadly consistent with the other two sources. The correlation between our aggregate flow measure and ICI flow is 0.68 at the monthly level and 0.80 at the quarterly level. The correlation between our measure and FoF flow is 0.55 at the quarterly level.

The differences in Figure A.1 between the three measures of aggregate flows likely reflect differences in mutual fund coverage. Specifically, the ICI flow tracks virtually all U.S. equity mutual funds that invest in both domestic and world equity markets.[37] The FoF flow, sourced from unpublished ICI data, aggregates unadjusted flows into and out of all U.S. mutual funds (including variable annuity long-term mutual funds). The FoF flow is calculated based

[37]The ICI is a trade association for the mutual fund industry, and virtually all U.S. mutual funds are ICI members (Warther, 1995).

on mutual fund assets in common stock, preferred stock, and rights and warrants.[38] In comparison, our mutual fund sample contains U.S. mutual funds that CRSP covers. CRSP collects historical data from various sources.[39] Due to the nature of the data collection process, CRSP's coverage is smaller than ICI's coverage.

D Additional Empirical Results

In this appendix, we provide additional empirical results.

Table A.1 presents the process of rotating MKT, SMB, and HML factors from their original counterparts. As depicted in the top-left panel, the MKT flow's volatility is approximately six times that of the SMB and HML flows. Although the pairwise correlation between these factors' flow is small, it is not zero. In the top-right panel, we present the correlations and standard deviations of the factors' fundamental returns $\mathbf{b}_k^\top \boldsymbol{\xi}_t$. These returns exhibit a high correlation, particularly between the MKT and HML factors. This high correlation arises from the discrepancy between the estimated SMB and HML portfolio weights $\mathbf{b}_k$ and the original Fama-French portfolio weights, which is discussed in Section 5.2.

The bottom panel shows the rotated MKT, SMB, and HML factors, which exhibit uncorrelated flows and fundamental returns. We also provide the portfolio weights of these rotated factors in the three figures at the end. Upon examination of these portfolio weights, we note that the rotated factors can still be interpreted as market, size, and value factors.

Table A.2 presents the IV first-stage regression results for the third column of Table 2. In this analysis, each factor's flow is regressed on its corresponding concurrent night return and the difference between one-month and six-month lagged flows. The flow into the MKT factor is found to significantly chase concurrent night returns. On the other hand, the flows into SMB and HML factors exhibit a more pronounced response to the differences in lagged

[38] See https://www.federalreserve.gov/apps/fof/SeriesAnalyzer.aspx?s=FA653064100&t=F.223&suf=Q.

[39] The sources include the Fund Scope Monthly Investment Company Magazine, the Investment Dealers Digest Mutual Fund Guide, Investor's Mutual Fund Guide, the United and Babson Mutual Fund Selector, and the Wiesenberger Investment Companies Annual Volumes.

Table A.1. Model rotation

original factors

	flow				**fundamental return**			
	correlation			std	correlation			std
	MKT	SMB	HML		MKT	SMB	HML	
MKT	1	0.11	0.25	5.34×10^{-2}	1	0.41	0.91	0.14
SMB	0.11	1	-0.11	0.83×10^{-2}	0.41	1	0.56	0.12
HML	0.25	-0.11	1	0.87×10^{-2}	0.91	0.56	1	0.28

rotated factors

	flow				**fundamental return**			
	correlation			std	correlation			std
	MKT	SMB	HML		MKT	SMB	HML	
MKT	1	0	0	0.85×10^{-2}	1	0	0	1
SMB	0	1	0	0.10×10^{-2}	0	1	0	1
HML	0	0	1	0.07×10^{-2}	0	0	1	1

	Low	2	3	4	High
Small	0.03	0.03	0.04	0.05	0.05
2	0.08	0.09	0.08	0.08	0.06
3	0.14	0.14	0.11	0.10	0.06
4	0.30	0.22	0.15	0.13	0.12
Big	1.80	0.97	0.71	0.60	0.44

rotated $\tilde{b}_{MKT}$

	Low	2	3	4	High
Small	0.30	0.28	0.36	0.38	0.37
2	0.65	0.66	0.61	0.56	0.37
3	0.89	0.85	0.68	0.58	0.39
4	0.70	0.59	0.31	0.29	0.27
Big	-5.24	-4.29	-2.23	-1.17	-0.20

rotated $\tilde{b}_{SMB}$

	Low	2	3	4	High
Small	-0.14	-0.06	-0.06	0.02	0.05
2	-0.39	-0.18	-0.10	-0.03	-0.03
3	-0.44	-0.02	0.15	0.21	0.19
4	-0.65	0.51	0.63	0.55	0.59
Big	-6.97	-0.82	1.27	3.36	1.96

rotated $\tilde{b}_{HML}$

Note: The top panel reports the correlations and standard derivations of flows and fundamental returns of the original MKT, SMB, and HML factors. The bottom panel reports the rotated MKT, SMB, and HML factors. The unit of flow is expressed as a percentage of the total stock market capitalization. The three figures at the end illustrate the portfolio weights of these rotated factors.

flows. The F-statistics hover around six for these factors.

In Table A.3, we perform a robustness check on Table 2 by focusing on unexpected flow components, which are measured following the methodology of Lou (2012). We first estimate each fund's Fama-French four-factor alphas using the fund's monthly returns and a one-year rolling window. These alphas are then used to predict fund flows in a regression model, and the residuals serve as unexpected flows. Next, we calculate stock-level unexpected flow-induced trading by aggregating fund-level unexpected flows based on lagged holdings. Finally, we repeat the analyses in Section 5 to construct unexpected flows for the Fama-

Table A.2. IV first-stage regression

	MKT flow	SMB flow	HML flow
concurrent night return	89.03×10^{-4}	-5.72×10^{-4}	-1.68×10^{-4}
	(2.35)	(-1.29)	(-0.49)
lag-1 flow − lag-6 flow	0.10	0.15	0.17
	(1.70)	(3.74)	(3.27)
constant	-5.01×10^{-4}	-0.19×10^{-4}	0.25×10^{-4}
	(-0.92)	(-0.29)	(0.59)
regression R^2	4.90%	4.81%	3.85%
F-statistics	6.19	6.06	4.81

Note: We report the IV first-stage regression results, with each factor's flow being regressed against its concurrent night return and the difference between one-month and six-month lagged flows. The figures in parentheses represent the t-statistics, computed using heteroskedasticity-robust standard errors.

Table A.3. Robustness check for second-stage regression using unexpected flows

	total return OLS	intraday return OLS	intraday return IV
λ_{MKT}	9.88	7.33	6.34
	(14.07)	(14.10)	(2.01)
λ_{SMB}	140.96	69.49	206.68
	(10.17)	(5.37)	(5.11)
λ_{HML}	116.70	55.69	219.49
	(3.06)	(1.71)	(1.34)

Note: In this table, we run the second-stage regression of 5×5 asset returns on the product of factor flows and the quantity of risk to estimate the price of flow-induced risk. We calculate flows based on unexpected components, following the methodology of Lou (2012). The unit of flow is expressed as a percentage of the total stock market capitalization, and the quantity of risk is expressed in terms of the annualized variance in returns. The first two columns display the OLS estimation results using total returns and intraday (open-to-close) returns. The third column outlines the IV estimation results using intraday returns, in which each factor flow is instrumented by the factor's concurrent overnight (close-to-open) return and the difference between one-month and half-year lagged flows. The figures in parentheses represent the t-statistics, computed using heteroskedasticity-robust standard errors.

French three factors and 5×5 test assets and then estimate the price of flow-induced risk λ_k. The findings, detailed in Table A.3, closely align with the baseline results in Table 2.

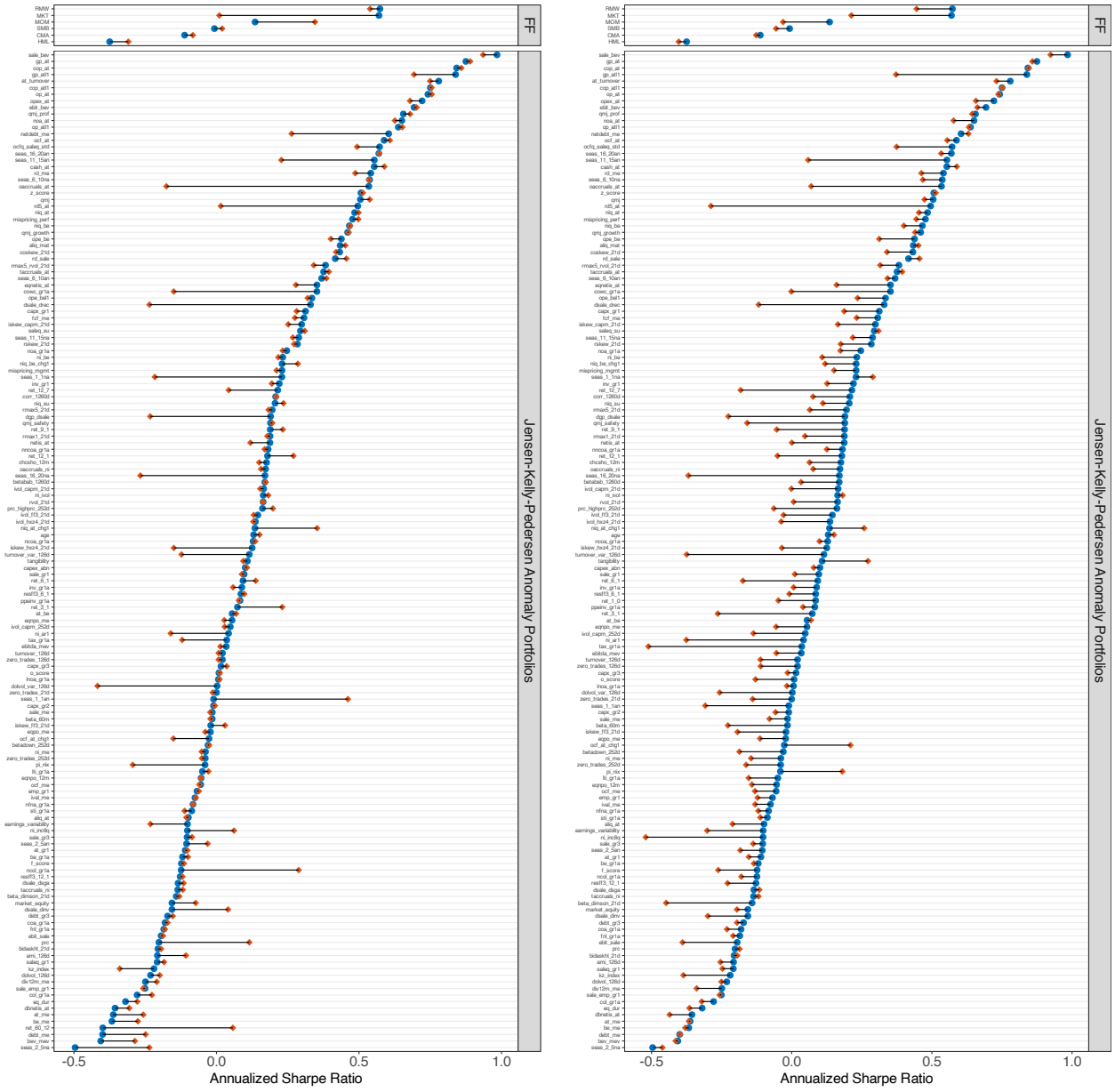
To rule out that the enhanced performance for existing anomalies is merely a mechanical effect of strategy combination, as per equation (19), Figure A.2 conducts a placebo test for Figure 6. This time, we substitute our MPIR strategy with the short-term reversal from

Jegadeesh (1990) and the long-term reversal from De Bondt and Thaler (1985). In each month, we scale the reversal strategy so that the combined portfolio allocates risk between the original anomaly and the reversal strategy proportional to their respective Sharpe ratios. Replacing (19), the new formula is

$$r^*_{j,t+1} = r_{j,t+1} + w_{j,t} r_{\text{rev},t+1} \text{ with } w_{j,t} := \max\left(\frac{\text{VOL}_j^{\{t\}}}{\text{SR}_j^{\{t\}}}, 0\right) \max\left(\frac{\text{SR}_{\text{rev}}^{\{t\}}}{\text{VOL}_{\text{rev}}^{\{t\}}}, 0\right). \tag{A.69}$$

Here, $\text{SR}_j^{\{t\}}$ and $\text{SR}_{\text{rev}}^{\{t\}}$ represent the Sharpe ratios of the anomaly portfolio j and reversal strategy, whereas $\text{VOL}_j^{\{t\}}$ and $\text{VOL}_{\text{rev}}^{\{t\}}$ represent their respective return standard deviations. Figure A.2 reveals negligible or negative improvements in mean and median Sharpe ratios: -0.02 and 0.00 for short-term reversal, and -0.11 and -0.14 for long-term reversal, respectively.

Figure A.2. Alternative reversal strategies and the Sharpe ratios of anomaly portfolios



(A) Short-term reversal **(B) Long-term reversal**

Note: The blue dots in the figure represent the out-of-sample annualized Sharpe ratio of the Jensen, Kelly, and Pedersen (2023) 154 portfolios, including the Fama-French-Carhart six factors and a large list of other firm characteristics-based anomaly portfolios. The red diamonds represent the Sharpe ratio of the portfolio that combines the original portfolio with two alternative reversal strategies: short-term reversal from Jegadeesh (1990) (panel A) and long-term reversal from De Bondt and Thaler (1985) (panel B). The expanding windows span from January 2000 to December 2004, and the out-of-sample testing period extends from January 2005 to September 2020.