Flow-Based Asset Pricing: Maximum Price Impact Ratio*

Yu An† Yinan Su‡ Chen Wang§

This version: July 27, 2022
First version: May 2, 2022

Abstract

We analyze the flow-driven fluctuations of the cross section of asset prices. We take the stance that price impacts of uninformative flows arise as marginal investors' risk compensation. We show that shorting the portfolio that incurs the maximum price impact for a given level of fundamental risk is the most efficient trading-against-flow strategy. To form this strategy, we build a new model of common factors of flows and common factors of fundamental returns and estimate the model using U.S. equity mutual fund flows into Fama-French three factors. Our strategy increases the out-of-sample annualized Sharpe ratio of 154 firm characteristics-based anomaly portfolios by an average of 0.3. Our evidence shows that risk-driven price impacts depend only on factor flows but not on idiosyncratic flows.

Keywords: asset pricing, cross section, flow, price impact, risk

JEL Codes: G12

*We thank Federico Bandi, Hank Bessembinder, Zhi Da, Darrell Duffie, Robin Greenwood, Ben Hébert, Arvind Krishnamurthy, Jiacui Li, Dong Lou, Paolo Pasquariello, Nagpurnanand Prabhala, Dongho Song, Zhaogang Song, Semih Üslü, Adrien Verdelhan, Zeyu Zheng, and seminar participants at Notre Dame and JHU Carey for helpful comments. All errors are our own.

†Carey Business School, Johns Hopkins University; yua@jhu.edu; https://sites.google.com/view/yu-an.
‡Carey Business School, Johns Hopkins University; ys@jhu.edu; https://www.suyinan.com/.
§Mendoza College of Business, University of Notre Dame; chen.wang@nd.edu; https://chenwang.one/.

1 Introduction

How do we understand asset price variations driven by uninformative and noisy trading flows? We take the stance that marginal investors bear extra fundamental risk when absorbing these uninformative flows, and price impacts arise as marginal investors' risk compensation. We formalize this intuition for the cross section of assets. We show that shorting the portfolio that incurs the maximum price impact for a given level of fundamental risk is the most efficient trading-against-flow strategy. To form this strategy, we build a new model of common factors of flows and common factors of fundamental returns and estimate the model using U.S. equity mutual fund flows into Fama and French (1993) factors. Our strategy increases the out-of-sample annualized Sharpe ratio of 154 firm characteristics-based anomaly portfolios by an average of 0.3. Our evidence shows that risk-driven price impacts depend only on factor flows but not on idiosyncratic flows. We complement the canonical risk-based asset pricing (e.g., Cochrane (2009) textbook) by adding flow to it.

Two-portfolio separation. We show that in any market with uninformative flows, the maximum Sharpe-ratio portfolio can be separated into two. The first portfolio uses only fundamental information to maximize the Sharpe ratio. Such fundamental investing has been extensively researched since Fama and French. The second portfolio maximizes the ratio of its price impact over fundamental risk, which is called the *maximum-price-impact-ratio (MPIR)* portfolio. Longing the fundamental-investing portfolio and shorting the MPIR portfolio maximize the overall Sharpe ratio. The two portfolios' risk exposures are proportional to their respective fundamental-based Sharpe ratio and flow-based price impact ratio.

The theory implies that, in order to use flow information to increase the Sharpe ratio, it suffices to estimate *only* the MPIR portfolio and combine this portfolio with fundamental investing. We next impose extra structures to estimate the MPIR portfolio and show that the MPIR strategy indeed improves fundamental-investing performance.

Factor model of price impacts. In order to find the MPIR portfolio, we need to tease

out the cross-sectional price impacts from the observed flows and returns of each asset. The following factor model reflects our risk view of price impacts,[1]

$$\text{price impact} = \sum_{\text{factor}} \text{factor price of risk} \quad \text{factor flow} \quad \text{quantity of risk exposure.}$$

Any asset's price impact depends on flows into a few risk factors. The quantity of risk exposure per dollar of factor flows is the covariance between the asset's and factor's fundamental returns. The factor's price of risk thus reflects marginal investors' aversion to absorbing each unit of factor risk and is subject to estimation. We next describe how to empirically obtain factor flow, quantity of risk exposure, and factor price of risk.

We use mutual fund flow-induced trading to empirically measure uninformative flows.[2] The idea is that mutual fund investors are mostly uninformed retail investors.[3] Their decision to buy or sell mutual fund shares reveals little information about the future fundamentals of underlying stocks. We convert mutual fund flows to stock-level flows, using the standard assumption that mutual funds buy or sell stocks in proportion to their lagged holdings. The resulting stock-level flows proxy the component of mutual fund flows that is induced by retail investors' activities, not by mutual fund managers' discretionary allocation. We use monthly U.S. equity mutual fund flows data from 2000 through 2020. Our test assets are the Fama-French 5 5 portfolios sorted on size and book-to-market equity.

In the first stage (1), we run a time-series regression of each test asset's flow $f_{n,t}$ on market (MKT), small-minus-big (SMB), and high-minus-low (HML) factor flows $q_{k,t}$,

$$\underbrace{f_{n,t}}_{\text{asset flow}} = \sum_{k \in \{\text{MKT,SMB,HML}\}} \underbrace{b_{n,k}}_{\text{flow beta}} \underbrace{q_{k,t}}_{\text{factor flow}} + \underbrace{e_{n,t}}_{\text{idiosyncratic flow}}. \tag{1}$$

The estimated flow betas are *ow mimicking portfolio weights*. Regression (1) implies that,

[1] An (2022) introduces flows to the Cochrane (2009) textbook asset-pricing framework and derives the factor model of price impacts by generalizing Ross (1976) arbitrage pricing theory.

[2] Mutual fund flow-induced trading has been used by various studies including Coval and Stafford (2007), Greenwood and Thesmar (2011), Lou (2012), Ben-David, Li, Rossi, and Song (2021a), and Li (2021).

[3] See, for example, Frazzini and Lamont (2008) and Ben-David, Li, Rossi, and Song (2021b).

holding all else equal, a one-dollar MKT-factor flow increases asset-n flow by $\$b_{n,\text{MKT}}$. Marginal investors are the rest of market who absorbs these flows, and our approach is agnostic about their identity. By market clearing, a one-dollar MKT-factor flow decreases marginal investors' holding in asset n by $\$b_{n,\text{MKT}}$. Therefore, the 25 flow betas $fb_{n,\text{MKT}}g$ mimic the *changes* in marginal investors' 5 5 holding caused by one dollar of MKT flow.

Turning to results of (1), the regression R^2 ranges from 50% for small companies to 80% for big companies, showing that flows into Fama-French factors explain large portions of common variation in asset flows. The 25 MKT-factor flow betas $b_{n,\text{MKT}}$ are remarkably close to the fraction of each asset's market capitalization to total stock market capitalization, consistent with Fama-French MKT weights. The SMB flow betas $b_{n,\text{SMB}}$ are positive for small companies and negative for big companies. The sum of all positive $b_{n,\text{SMB}}$ is close to 1 and the sum of all negative $b_{n,\text{SMB}}$ is close to 1, consistent with Fama-French SMB weights. For HML, only the BL and B2 test assets (Biggest companies with Lowest and 2nd lowest book-to-market equity) have negative flow betas, and all other assets have positive flow betas. Our flow betas differ from Fama-French HML weights, because BL and B2 assets combined have over 50% of the total stock market capitalization. The total quantity of flows in the HML direction is really BL and B2 assets against all other assets. Overall, the first-stage evidence supports two results: (i) Fama-French factor flows are systematic flows and (ii) absorbing Fama-French factor flows exposes marginal investors to systematic risks. These two results ensure that using only Fama-French factors in the second-stage regression is sufficient to capture a large portion of cross-sectional variation in price impacts.[4]

In the second stage (2), we run a panel regression of asset returns $r_{n,t}$ on the product of factor flows $q_{l,t}$ and quantity of risk $\text{cov}(\mathbf{R}_0(n), \mathbf{b}_k^{\succ}\mathbf{R}_0)$,

$$\underbrace{r_{n,t}}_{\text{asset return}} = \sum_{k,l2f\text{MKT,SMB,HML}g} \underbrace{\lambda_{k,l}}_{\text{price of risk}} \underbrace{q_{l,t}}_{\text{factor flow}} \underbrace{\text{cov}(\mathbf{R}_0(n), \mathbf{b}_k^{\succ}\mathbf{R}_0)}_{\text{quantity of risk}} + \xi_{n,t}. \quad (2)$$

[4]This claim is true under a bounded maximum price impact ratio, in a similar spirit to the good-deal bound of Cochrane and Saa-Requejo (2000). See An (2022) for precise theoretical statement and proof.

Importantly, we use flow betas to form mimicking portfolio $\mathbf{b}_k = (b_{1,k}, b_{2,k}, \ldots, b_{25,k})^\top$ for each factor k. In equilibrium, marginal investors trade against this portfolio $\mathbf{b}_k$ for every dollar of factor flow $q_{k,t}$. We use one-year return[5] to proxy the fundamental return $\mathbf{R}_0$. For each asset n, the quantity of risk induced by every dollar of factor-k flow is $\text{cov}(\mathbf{R}_0(n), \mathbf{b}_k^\top \mathbf{R}_0)$.

The quantity of risk differentiates our regression (2) from the inverse demand elasticity, which directly regresses returns on flows. In our framework, because flows change marginal investors' holdings and fundamental returns are risky, the marginal investors are averse to the changes in the fundamental *risk* exposures induced by flows. The marginal investors are not averse to flows per se.[6] The quantity of risk reflects our risk stance of price impacts. Note that two factors l and k could have cross impacts on each other, so the estimated price of risk $\lambda_{k,l}$ measures how factor-l flow impacts factor-k price. We impose a theory-founded structural restriction when estimating (2), such that the 3 3 prices of risk $\lambda_{k,l}$ have 3 degrees of freedom. Economically, 3 factors should have 3 prices of risk, not 9 prices of risk.

Turning to results of (2), we find that the risk induced by SMB flows has the greatest explanatory power for the cross-sectional variation of price impacts. The MKT and HML factors have roughly the same explanatory power. This finding may be surprising given that MKT flow is about six times as volatile as SMB and HML flows. This is because the SMB and HML factors have a much higher price of risk than the MKT factor, and the quantity of risk induced by every dollar of flow also differs. Not all flows are created equal—the price of risk and the quantity of risk matter for the cross section of price impacts.

The main endogeneity concern of regression (2) is that flows $q_{l,t}$ chase concurrent returns $r_{n,t}$, leading to potential over-estimation of the price of risk. If we correctly estimate the price of risk, the estimated price impacts should revert 100% in the future; if we over-estimate the price of risk, the price impacts should revert less than 100%. Therefore, to address the endogeneity concern, we estimate the ratio of reversion to price impact in sample, use this

[5] One-year returns are the most commonly used long-horizon returns in the asset pricing literature. See Koijen and van Nieuwerburgh (2011) for a review.

[6] Our channel of price impacts is therefore different from the noise trader risk of De Long, Shleifer, Summers, and Waldmann (1990).

ratio to shrink the estimated prices of risk, and use the shrunk prices of risk to construct out-of-sample MPIR strategy—our ultimate objective. Besides this parsimonious shrinkage, we also conduct alternative OLS and IV estimations to ensure that the main specification captures mostly price impacts.

Common versus idiosyncratic flows. Our model implies that risk-driven price impacts depend only on factor flows but not on idiosyncratic flows (the residuals $e_{n,t}$ from (1)). How do we empirically test this restriction? We define *anomaly* price impacts as the price impacts created by idiosyncratic flows. We construct two MPIR portfolios, one using only the factor flows as in (2), and the other using the factor plus anomaly price impacts. We test the difference between these two MPIR portfolios. This intuitive diagnostic procedure generalizes Gibbons, Ross, and Shanken (1989) to flows—we prove that the difference in the squared MPIR equals the joint χ^2 test statistic for anomaly price impacts.

Turning to empirical results, the price dislocation of the factor-based MPIR portfolio reverts to zero in six months on average. This reversion is consistent with our stance that the price impacts of Fama-French factor flows are caused by marginal investors' risk aversion. The factor-plus-idiosyncratic-flow MPIR portfolio has a much larger price dislocation. However, the dislocation barely reverts even after one year, showing that the anomaly price impacts are not driven by the risk channel. The larger price dislocation estimated in sample is likely caused by over-fitting, similar to the in-sample Markowitz optimization. The sharp contrast in the price reversion shows that risk-driven price impacts depend only on factor flows but not on idiosyncratic flows. From a practical perspective, trading against idiosyncratic flows is unreliable, and forming a structured factor model is valuable.

Out-of-sample trading strategy. Having established the factor model of price impacts, we construct the model-implied optimal trading-against-flow strategy using the MPIR portfolios. The improvements of investment outcomes on top of firm characteristics-based anomaly portfolios are wide-ranging and economically significant. Of the 154 Fama-French and Jensen, Kelly, and Pedersen (2021) portfolios, 143 or 93% receive a positive increase in

Sharpe ratio. The average change in annualized Sharpe ratio is 0.3.

The strategy is formed strictly out-of-sample using ex-ante available mutual fund flow signals and in-sample estimated parameters. Specifically, at the start of month $t+1$, we form a staggered strategy that trades against the MPIR portfolios of the past six months in equal weights. These MPIR portfolios incur the greatest price dislocation per unit of risk in months t 5 to t, and the staggered strategy is to take advantage of their subsequent price reversion. Our MPIR strategy is effectively a monthly rebalancing portfolio on the Fama-French 5 5 test assets, by trading against the MKT, SMB, and HML flows into assets. The trading-against-flow strategy is added to each of the 154 anomaly portfolios to form the optimal investment (with a theory-implied mixing ratio). The results reported above are acquired with the benchmark setting where the expanding estimation windows start with January 2000 to December 2004 and the out-of-sample testing period is January 2005 to September 2020. The results are robust to a series of alternative specifications, including changing skip and lookback periods and alternative estimation of the price of risk.

Besides the practical relevance, the out-of-sample investment performance of the MPIR strategy implies that the MPIR is sizable in real life (the annualized price impact ratio of our strategy is 0.5). Whereas the maximum Sharpe ratio equals the minimum volatility of stochastic discount factor (SDF) times the risk-free rate, the MPIR equals the minimum volatility of *ow-induced changes* in SDF.[7] Our results therefore imply that flow-induced price impacts can explain a significant portion of SDF variation (the excess volatility puzzle).

Related literature. From a theoretical perspective, the two-portfolio separation connects the Fama-French-style asset pricing that studies expected returns with demand-system asset pricing (Koijen and Yogo (2019)) and microstructure literature that studies price impacts.[8] Our objective for studying price impacts differs significantly from the existing litera-

[7]The first result is the Hansen and Jagannathan (1991) bound. The second result is proved in An (2022).

[8]For related literature, see, for example, Hasbrouck (1988), Hasbrouck (1991), Lo and Wang (2000), Hasbrouck and Seppi (2001), Greenwood (2005), Coval and Stafford (2007), Cremers and Mei (2007), Andrade, Chang, and Seasholes (2008), Greenwood and Thesmar (2011), Lou (2012), Boulatov, Hendershott, and Livdan (2013), Chang, Hong, and Liskovich (2015), Pasquariello and Vega (2015), Bouchaud, Bonart, Donier, and Gould (2018), Da, Larrain, Sialm, and Tessada (2018), Frazzini, Israel, and Moskowitz (2018),

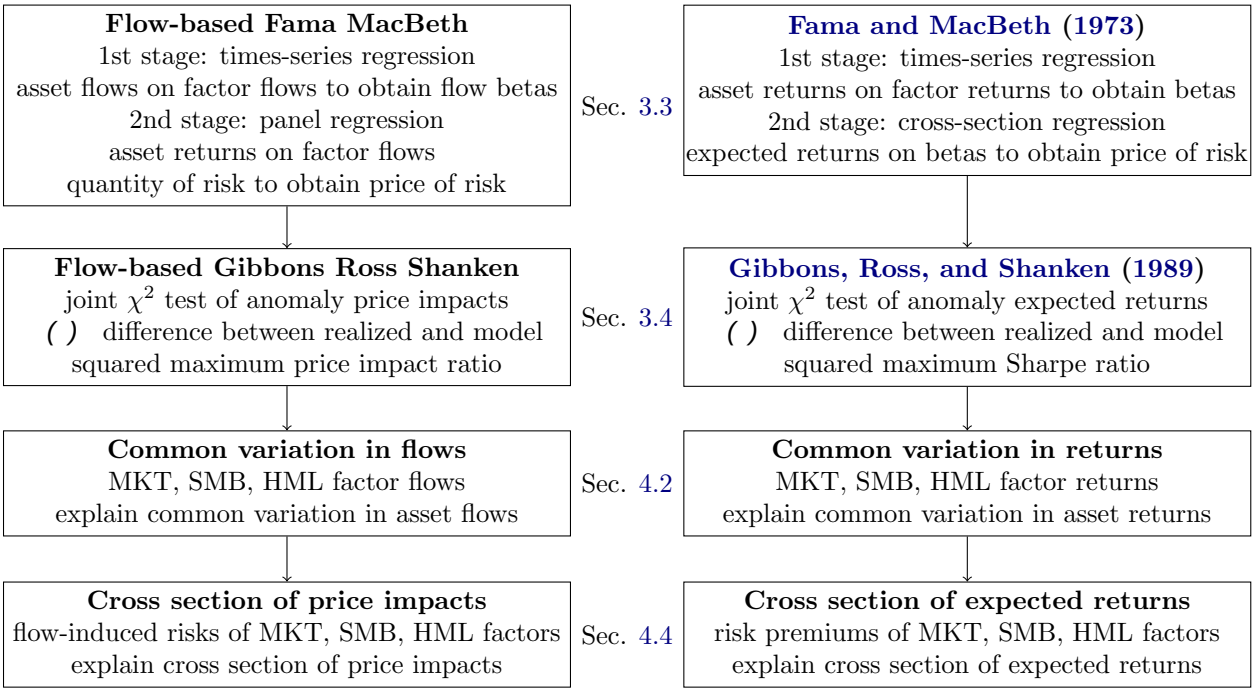
ture. We develop the new factor model to search for the MPIR portfolio in the cross section, rather than estimating the price impacts of any particular assets or factors. For the purpose of using flow information to increase the Sharpe ratio, it suffices to estimate *only* the MPIR portfolio and combine this portfolio with fundamental investing.

From a methodological perspective, we show that the classic factor models can be applied to study not only expected returns but also price impacts. Figure 1 presents our paper's analogy to standard asset pricing (Fama and MacBeth (1973); Gibbons, Ross, and Shanken (1989); Fama and French (1993)). Whereas the maximum Sharpe ratio is central to the standard risk-based asset pricing, the MPIR is central to our flow-induced-risk-based asset pricing. Theoretically and empirically, we show that MPIR improves fundamental-investing performance. Modern approaches to factor models, especially those applying machine-learning techniques, could potentially be applied to our factor model of price impacts (see Giglio, Kelly, and Xiu (2021) and Nagel (2021) for overviews). Our factor-model approach to price impacts is motivated by Kozak, Nagel, and Santosh (2018). Our framework formalizes the intuitions of how common flows generate price impacts for the cross section of assets that have common risks, and shows how to estimate such impacts in the data.

From an empirical perspective, we use mutual fund flow-induced trading of Coval and Stafford (2007) and Lou (2012) to proxy uninformative flows. We address endogeneity concern by shrinking the estimated price of risk by the ratio of reversion to price impacts. This shrinkage idea is motivated by permanent versus transient price impacts from the microstructure literature (O'hara (1997)). We use the in-sample shrunk price of risk to construct out-of-sample MPIR strategy. By not relying on particular instruments, our estimation is easily extendable to alternative factors and flow data other than mutual funds.

Li, Pearson, and Zhang (2020), Schmickler (2020), Ben-David, Li, Rossi, and Song (2021a), Huang, Song, and Xiang (2021), Li (2021), Koijen, Richmond, and Yogo (2021), Peng and Wang (2021), Dong, Kang, and Peress (2022a), Dong, Krystyniak, and Peng (2022b), Gabaix and Koijen (2022), Haddad, Huebner, and Loualiche (2022), Li and Lin (2022), and Pavlova and Sikorskaya (2022).

Figure 1. Flow-based asset pricing and analogy with standard asset pricing



2 Theory: Price Impact Ratio and Two-Portfolio Separation

We present a simple yet general setup, where flows generate concurrent price impacts but do not affect long-term asset payoffs. By extending the textbook portfolio calculations to this setup with flows, we obtains the theoretical backbone of the paper—price impact ratio and two-portfolio separation theorem.[9]

Asset $n = 1, 2, \ldots, N$ has a flow f_n at time 0, where $\mathbf{f} = (f_1, f_2, \ldots, f_N)^\top$. Asset n has a payoff X_n at time 1, where $\mathbf{X} = (X_1, X_2, \ldots, X_N)^\top$. Without loss of generality, we assume that $\text{var}(\mathbf{X})$ has full rank.[10] Flows are uninformative about payoffs. That is, payoffs $\mathbf{X}$ do not depend on $\mathbf{f}$. The gross risk-free rate between time 0 and 1 is a constant R_F. Under flow $\mathbf{f}$, we denote the time-0 price of asset n as $P_n(\mathbf{f})$. Under flow $\mathbf{f}$, asset returns are defined as

$$\mathbf{R}(\mathbf{f}) = \left(\frac{X_1}{P_1(\mathbf{f})}, \frac{X_2}{P_2(\mathbf{f})}, \ldots, \frac{X_N}{P_N(\mathbf{f})} \right)^\top. \tag{3}$$

[9] We use bold font notation for matrices and vectors. We use $\mathbf{A}^\top$ to denote the transpose of matrix $\mathbf{A}$. We use $\mathsf{E}[\mathbf{b}]$ and $\text{var}(\mathbf{b})$ to denote the expectation and variance-covariance matrix of a random vector $\mathbf{b}$.

[10] Otherwise, one can rotate the payoff $\mathbf{X}$ and flow $\mathbf{f}$ to linearly independent portfolios.

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

Price impacts are defined as the time-0 percentage price change with and without flow $\mathbf{f}$,

$$\Delta\mathbf{p}(\mathbf{f}) = \left(\frac{P_1(\mathbf{f}) \quad P_1(\mathbf{0})}{P_1(\mathbf{0})}, \frac{P_2(\mathbf{f}) \quad P_2(\mathbf{0})}{P_2(\mathbf{0})}, \ldots, \frac{P_N(\mathbf{f}) \quad P_N(\mathbf{0})}{P_N(\mathbf{0})} \right)^{>}. \tag{4}$$

For any portfolio $\mathbf{c} = (c_1, c_2, \ldots, c_N)^{>}$, where c_n is the dollar amount invested in asset n, we define the *price impact ratio* of portfolio $\mathbf{c}$ in the economy with flow $\mathbf{f}$ as

$$\mathrm{PIR}(\mathbf{c}, \mathbf{f}) := \frac{\mathbf{c}^{>}\Delta\mathbf{p}(\mathbf{f})}{\sigma(\mathbf{c}^{>}\mathbf{R}(\mathbf{0}))}, \tag{5}$$

where $\mathbf{c}^{>}\Delta\mathbf{p}(\mathbf{f})$ is the portfolio's price impact and $\sigma(\mathbf{c}^{>}\mathbf{R}(\mathbf{0}))$ is the portfolio's fundamental-return risk. Appendix A.1 proves the following proposition that the price impact ratio roughly equals the *ow-induced changes* in Sharpe ratios.

PROPOSITION 1. *We have*

$$R_F \mathrm{PIR}(\mathbf{c}, \mathbf{f}) = \mathrm{SR}(\mathbf{c}, \mathbf{0}) \quad \mathrm{SR}(\mathbf{c}, \mathbf{f}), \tag{6}$$

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

$$\mathrm{SR}(\mathbf{c}, \mathbf{f}) := \frac{\mathsf{E}[(\mathbf{W}(\mathbf{f})\mathbf{c})^{>}(\mathbf{R}(\mathbf{f}) \quad R_F\mathbf{1})]}{\sigma((\mathbf{W}(\mathbf{f})\mathbf{c})^{>}\mathbf{R}(\mathbf{f}))}, \tag{7}$$

with the $N \quad N$ *diagonal matrix* $\mathbf{W}(\mathbf{f}) := \mathrm{diag}(P_1(\mathbf{f})/P_1(\mathbf{0}), P_2(\mathbf{f})/P_2(\mathbf{0}), \ldots, P_N(\mathbf{f})/P_N(\mathbf{0}))$. *Because asset prices change from* $P_n(\mathbf{0})$ *to* $P_n(\mathbf{f})$*, the dollar amount invested in asset* n *changes from* c_n *to* $c_n P_n(\mathbf{f})/P_n(\mathbf{0})$*, so* $\mathbf{W}(\mathbf{f})\mathbf{c}$ *is the portfolio weights under ow* $\mathbf{f}$*.*

For example, suppose that a portfolio $\mathbf{c}$ has a 5% expected return. After flows $\mathbf{f}$ have arrived, the portfolio's price increases by 1% today, so the expected return decreases to approximately 4%. Equation (6) formalizes this accounting relationship $1\% = 5\% \quad 4\%$.

Next, we define the *maximum-price-impact-ratio (MPIR)* portfolio under flow $\mathbf{f}$ as

$$\tilde{\mathbf{c}}\ (\mathbf{f}) := \mathrm{var}(\mathbf{R}(\mathbf{0}))\ {}^{1}\Delta\mathbf{p}(\mathbf{f})\ 2\ \arg\max_{\mathbf{c}2\mathsf{R}^N} \mathrm{PIR}(\mathbf{c},\mathbf{f}), \tag{8}$$

and the maximum-Sharpe-ratio portfolio under flow $\mathbf{f}$ as

$$\mathbf{c}\ (\mathbf{f}) := \mathbf{W}(\mathbf{f})\ {}^{1}\mathrm{var}(\mathbf{R}(\mathbf{f}))\ {}^{1}\mathsf{E}[\mathbf{R}(\mathbf{f}) \quad R_F\mathbf{1}]\ 2\ \arg\max_{\mathbf{c}2\mathsf{R}^N} \mathrm{SR}(\mathbf{c},\mathbf{f}). \tag{9}$$

We do not make any specific assumptions on distributions and utility functions, or impose any factor-model and equilibrium structures. Appendix A.2 proves the following theorem.

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

$$\underbrace{\mathbf{c}\ (\mathbf{f})}_{\substack{\text{max. Sharpe ratio port. with flow}}} = \overbrace{\mathbf{c}\ (\mathbf{0})}^{\substack{\text{max. Sharpe ratio port. without flow}}} \quad R_F\ \underbrace{\tilde{\mathbf{c}}\ (\mathbf{f})}_{\substack{\text{max. price impact ratio port.}}}. \tag{10}$$

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

Equation (10) 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 $\mathbf{R}(\mathbf{0})$. That is, the fundamental-investing portfolio $\mathbf{c}\ (\mathbf{0})$ completely ignores the flow information. The second portfolio $\tilde{\mathbf{c}}\ (\mathbf{f})$ maximizes the price impact ratio under flow $\mathbf{f}$. Shorting the MPIR portfolio is the most efficient strategy to trade against uninformative flows, or equivalently, to provide liquidity to flows.

Longing the fundamental-investing portfolio $\mathbf{c}\ (\mathbf{0})$ and shorting the MPIR portfolio $\tilde{\mathbf{c}}\ (\mathbf{f})$ maximize the overall Sharpe ratio. 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. This is an approximation because of the multiplication

by the risk-free rate R_F in (10). The precise risk allocation is proportional to the volatility of the SDF $M(\mathbf{0})$ in the economy without flow and the volatility of the *changes* of the SDF $M(\mathbf{f}) \quad M(\mathbf{0})$ with and without flow.[11] In any reasonable model, the price impact ratio is increasing in flows.[12] In periods with larger uninformative 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.

Empirically, Theorem 1 implies that in order to use flow information to increase the Sharpe ratio, it suffices to estimate *only* the MPIR portfolio and then combine this portfolio with fundamental investing. We show this empirical procedure does work.

Theorem 1 implies that any Sharpe-ratio-maximizing investor should simultaneously be a fundamental investor and a liquidity provider who trades against flows. Excessive specialization in one role may reflect Sharpe ratios left on the table. For example, Koijen and Yogo (2019) find that institutions vary drastically in demand elasticity, suggesting their tendency to specialize. Our trading-against-MPIR strategy increases the out-of-sample annualized Sharpe ratio of 154 firm characteristics-based anomaly portfolios by an average of 0.3.

Relatedly, Duffie (2012) argues that it is hard in practice to eliminate the proprietary-trading incentive of market-making desks at bank dealers, which is a stated objective of the Volcker rule. Theorem 1 suggests one potential solution by changing the performance evaluation measure of the market-making desks from the Sharpe ratio to the price impact ratio. Price-impact-ratio maximization incentivizes risk-constrained market makers to provide liquidity to portfolios that suffer the largest flow-induced price dislocation in the cross section. The price impact ratio can be concretely measured in the data using our factor model.

The separation between fundamental investing and liquidity provision arises because flows $\mathbf{f}$ are uninformative of fundamental returns $\mathbf{R}_0$. This implies that uninformative-flow-driven return fluctuations are independent of fundamental-driven fluctuations. We model these

[11] Hansen and Jagannathan (1991) prove that the volatility of the SDF $M(\mathbf{0})$ equals the maximum Sharpe ratio over R_F. An (2022) proves that the volatility of the change of the SDF $M(\mathbf{f}) \quad M(\mathbf{0})$ equals the MPIR.

[12] For example, in any linear price impact model, the price impact ratio is proportional to flows.

two sources of return fluctuations in a simple and meaningful setup. The economies with and without flows correspond to two states of the world at time 0, and the fundamental payoffs realize at time 1. Therefore, flow-driven return fluctuations happen at time 0 and fundamental-driven fluctuations happen between time 0 and 1. This setup makes sense because the short-term price impacts created by uninformative flows take a long time to revert until fundamental payoffs realize. Modeling flows as two states of the world at time 0, instead of adding an extra time period, reveals the economics of short-term price impacts versus long-term fundamental payoffs and achieves the two-portfolio separation.

3 Empirical Methods

In order to find the MPIR portfolio, we need to tease out the cross-sectional price impacts from the observed flows and returns of each asset. To this end, we introduce the factor model of price impacts, show how to construct portfolio flows and estimate the model, and test whether price impacts depend only on factor flows.

3.1 Factor Model of Price Impacts

We now specify how flows generate price impacts for the cross section of assets. The model's key feature is that any asset's price impact depends only on flows into a few common risk factors and the asset's risk exposure to each common factor. The model is hence called the "factor model of price impacts." Our cross-sectional framework contrasts with an isolated framework where each asset's price impact depends only on its own flow.

Specifically, the model has K factors. The flow into each factor k in dollar amounts is denoted as q_k. The portfolio weights of factor k are $\mathbf{b}_k = (b_{1,k}, b_{2,k}, \ldots, b_{N,k})^\top$. We use $\mathbf{R}_0$ to denote the N assets' fundamental returns, as defined in (3). Therefore, asset n's fundamental

risk exposure to factor k is $\text{cov}(\mathbf{R}_0(n), \mathbf{b}_k^{\succ}\mathbf{R}_0)$. The N assets' price impacts $\Delta\mathbf{p}(\mathbf{f})$ satisfy

$$\Delta\mathbf{p}(\mathbf{f}) = \sum_{k=1}^{K}\sum_{l=1}^{K} \lambda_{k,l} q_l \text{cov}(\mathbf{R}_0, \mathbf{b}_k^{\succ}\mathbf{R}_0). \tag{11}$$

Because $\text{cov}(\mathbf{R}_0, \mathbf{b}_k^{\succ}\mathbf{R}_0)$ is the quantity of risk per dollar of factor-k flow and q_l is the dollar-amount flow into factor l, the parameter $\lambda_{k,l}$ has a price-of-risk meaning. Specifically, $\lambda_{k,l}$ measures how flow q_l into factor l impacts the price of factor k.

This intuitive model generalizes the canonical factor model of expected returns.[13] Whereas the canonical model concerns expected returns, our model concerns risk-driven price impacts, which are *ow-induced changes* in expected returns. This is why when substituting the factor flow q_l in (11) by a constant one, we effectively get back the canonical model.

A complication of (11) is a structural restriction on $\lambda_{k,l}$. The K factors could have correlated fundamental returns $\mathbf{b}_k^{\succ}\mathbf{R}_0$ and correlated flows q_k, so they should have some cross impacts on each other, i.e., $\lambda_{k,l} \neq 0$ for $k \neq l$. At the same time, the K K prices of risk $\lambda_{k,l}$ should not be K^2 free parameters. After all, one risk factor should have only one price of risk. Therefore, intuitively, one should impose some restrictions on how price impacts behave in the cross section. This intuition is formalized via an Irrelevance of Uncorrelated Flows (IUF) restriction. Specifically, we use the K original factors to form K orthogonalized factors that have uncorrelated fundamental returns $\tilde{\mathbf{b}}_k^{\succ}\mathbf{R}_0$ and uncorrelated flows $\tilde{q}_k$. The IUF assumes that these orthogonalized factors have no cross impacts and each factor has its own price of risk. As a result, model (11) simplifies to

$$\Delta\mathbf{p}(\mathbf{f}) = \sum_{k=1}^{K} \tilde{\lambda}_k \tilde{q}_k \text{cov}(\mathbf{R}_0, \tilde{\mathbf{b}}_k^{\succ}\mathbf{R}_0). \tag{12}$$

Differing from (11), the K orthogonalized factors have in total K prices of risk in (12). Because of the IUF restriction, we will first orthogonalize the original model (11) to (12),

[13] An (2022) provides the theory foundation of the factor model of price impacts by generalizing Cochrane (2009) textbook setup and Ross (1976) APT to model price impacts and flows.

regress for prices of risk $\tilde{\lambda}_k$ based on (12), and then rotate back to (11) to obtain $\lambda_{k,l}$.

3.2 Constructing Portfolio Flows

To estimate the model (11), we need to choose N test assets and K factors, which could both be portfolios of M underlying stocks.[14] Empirically, we will choose M 3,000 stocks, $N = 25$ test assets, and $K = 3$ factors. We now show how to construct portfolio flows.

We observe the flow $f_{m,t}^{\{\text{stock}\}}$ into M stocks at time t. We have K portfolios, where the weight of stock m in portfolio k is denoted as $w_{m,k}$. We write the portfolio weights $w_{m,k}$ as an M K matrix $\mathbf{W}$. The flow $q_{k,t}$ in portfolio k at time t is constructed as

$$\begin{pmatrix} q_{1,t} \\ q_{2,t} \\ \\ \\ q_{K,t} \end{pmatrix} = (\mathbf{W}^{>}\mathbf{W})^{\ 1}\mathbf{W}^{>} \begin{pmatrix} f_{1,t}^{\{\text{stock}\}} \\ f_{2,t}^{\{\text{stock}\}} \\ \\ \\ f_{M,t}^{\{\text{stock}\}} \end{pmatrix}. \tag{13}$$

The key to understand (13) is that portfolio weight $w_{m,k}$ is the rule for allocating portfolio flows to individual stocks. If portfolio-k flow increases by \$1, stock-m flow increases by $\$w_{m,k}$. Therefore, if the K portfolio flows are $q_{1,t}, q_{2,t}, \ldots, q_{K,t}$, the resulting flow into stock m is $\sum_{k=1}^{K} w_{m,k} q_{k,t}$. Our goal is to find the portfolio flows $q_{1,t}, q_{2,t}, \ldots, q_{K,t}$ that best approximate the observed stock-level flows $f_{1,t}^{\{\text{stock}\}}, f_{2,t}^{\{\text{stock}\}}, \ldots, f_{M,t}^{\{\text{stock}\}}$. This approximation is achieved with a cross-sectional regression of asset flows $f_{n,t}$ on portfolio weights $w_{n,1}, w_{n,2}, \ldots, w_{n,K}$, which obtains the pseudoinverse $(\mathbf{W}^{>}\mathbf{W})^{\ 1}\mathbf{W}^{>}$ in (13).

The portfolio-flow construction differs from the portfolio-return construction, where one multiplies the transpose $\mathbf{W}^{>}$ with stock returns. Intuitively, portfolio flows first go into portfolios and are then allocated to individual stocks according to portfolio weights. Therefore, one should use the pseudoinverse $(\mathbf{W}^{>}\mathbf{W})^{\ 1}\mathbf{W}^{>}$ to construct in the reverse direction from

[14] Our framework is not restricted to stocks. We use "stocks" simply because we later apply to stocks.

stock flows to portfolio flows.[15] Appendix B.1 also gives two simple numerical examples to illustrate that the pseudoinverse $(\mathbf{W}^{\top}\mathbf{W})^{-1}\mathbf{W}^{\top}$, rather the transpose $\mathbf{W}^{\top}$, leads to portfolio flows that agree with common intuitions.

3.3 Flow-Based Fama-MacBeth Regression

We show how to estimate price impact model (11) to tease out the cross-sectional price impacts from the observed asset flows and returns.

Our data inputs are $\{r_{n,t}\}$ and $\{f_{n,t}\}$, the return and flow of test asset n at time t for $n = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. We also have the time series of K factor flows $\{q_{k,t}\}$ for $k = 1, 2, \ldots, K$ and $t = 1, 2, \ldots, T$. Because we focus on the variation of price impacts, we remove the unconditional time-series mean of return $r_{n,t}$ and flow $f_{n,t}$ and $q_{k,t}$ for each test asset and factor, so that we do not carry extra constants in subsequent regressions.

We assume that the price impact model (11) repeats over time. We do not observe the price impact component per se, but observe the test assets' return, which equals the price impact plus fluctuation driven by fundamental information. Specifically, we assume that

$$r_{n,t} = \delta_{n,t} + \xi_{n,t} = \sum_{k=1}^{K}\sum_{l=1}^{K} \lambda_{k,l} q_{l,t} \mathrm{cov}(\mathbf{R}_0(n), \mathbf{b}_k^{\top}\mathbf{R}_0) + \xi_{n,t}, \tag{14}$$

where the price impact component $\delta_{n,t}$ of asset n at time t depends on the concurrent factor flows $q_{1,t}, q_{2,t}, \ldots, q_{K,t}$. The $K \times K$ prices of risk $\lambda_{k,l}$ satisfy the IUF restriction. The residual $\xi_{n,t}$ is the fundamental return component.

To tease out the price impact component $\delta_{n,t}$ from the total return $r_{n,t}$, we assume that $\mathsf{E}[q_{k,t}\xi_{n,t}] = 0$ for all factors $k = 1, 2, \ldots, K$ and assets $n = 1, 2, \ldots, N$. That is, we require factor flows $q_{k,t}$ to be uninformative about assets' concurrent fundamental returns $\xi_{n,t}$. We address the endogeneity concern of $q_{k,t}$ in Section 4.4.

[15] Section 2.2 of An (2022) provides the theory foundation for portfolio-flow construction. Specifically, price impacts arise because marginal investors absorb the aggregate payoff risk induced by flows, which is an inner product between flows and returns. To keep the aggregate payoff risk invariant, portfolio-flow construction needs to run in the opposite direction of portfolio-return construction.

We estimate fundamental-risk matrix var($\mathbf{R}_0$) and take it as a known model input.[16] To robustly estimate var($\mathbf{R}_0$) that is mostly free from the confounding short-term price impacts, we empirically use long-horizon returns.

First-stage regression. First, we estimate the portfolio weights $\mathbf{b}_k$ in (14) that mimic factor k using the N test assets. For each asset $n = 1, 2, \ldots, N$, we run a time-series regression of asset flow $f_{n,t}$ on factor flows $q_{1,t}, q_{2,t}, \ldots, q_{K,t}$,

$$f_{n,t} = \sum_{k=1}^{K} b_{n,k} q_{k,t} + e_{n,t}. \tag{15}$$

This time-series regression recovers the beta $b_{n,k}$ of asset-n flow to factor-k flow. The key insight is that flow betas $b_{n,k}$ ($n = 1, 2, \ldots, N$) are mimicking portfolio weights for factor flow $q_{k,t}$. Regression (15) implies that, holding all else equal, a one-dollar factor-k flow increases asset-n flow by \$$b_{n,k}$. By market clearing, a one-dollar factor-k flow decreases marginal investors' holding in asset n by \$$b_{n,k}$. Therefore, the flow betas $\mathbf{b}_k = (b_{1,k}, b_{2,k}, \ldots, b_{N,k})^\top$ mimic the *changes* in marginal investors' asset holding caused by one dollar of factor-k flow.

Our flow mimicking portfolios share an analogy with the Fama-MacBeth return mimicking portfolios. Applying equation (13), the time-t flows into our mimicking portfolios $\mathbf{B} = (b_{n,k})$ are $(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{f}_t$, where the flows into the N test assets are $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \ldots, f_{N,t})^\top$. In the standard Fama-MacBeth framework, the mimicking portfolios' returns are $(\boldsymbol{\beta}^\top\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^\top\mathbf{r}_t$, where $\boldsymbol{\beta}$ are return betas and $\mathbf{r}_t$ are the returns of the N test assets. Between the flow and return frameworks, the two differences in constructing portfolio flow/return and in the economic interpretation of flow/return betas exactly cancel out.

We stress that the mimicking portfolio weights $\mathbf{B} = (b_{n,k})$ and the original portfolio weights $\mathbf{W} = (w_{m,k})$ that construct factor flows in (13) are distinct objects. They do not have the same dimension—$\mathbf{B}$ is the weights on the N test assets and $\mathbf{W}$ is on the M stocks.

That being said, one may still conjecture that $\mathbf{B}$ should somehow resemble the original

[16] If we assume additionally that the cross-sectional fundamental returns $\boldsymbol{\xi}_t = (\xi_{1,t}, \xi_{2,t}, \ldots, \xi_{N,t})^\top$ is i.i.d. over time (as in Theorem 2), the variance-covariance matrix $\boldsymbol{\Sigma}_\xi$ of fundamental returns $\boldsymbol{\xi}_t$ equals var($\mathbf{R}_0$).

portfolio weights $\mathbf{W}$, if we have chosen a good set of factors to start with. For example, the flow betas (mimicking portfolio weights) for the SMB factor should long the test assets of small companies and short the test assets of big companies. The following proposition formalizes this intuition. Appendix A.3 provides a proof.

PROPOSITION 2. *When test assets are all the stocks ($N = M$), ow betas recover original portfolio weights $\mathbf{B} = \mathbf{W}$ if and only if the factor ows $q_{1,t}, q_{2,t}, \ldots, q_{K,t}$ can be rotated into K principal components of asset ows $f_{n,t}$.*

This proposition shows that the flow betas $\mathbf{B}$ do not have to resemble the original portfolio weights $\mathbf{W}$. But if they do, we then know that we have chosen a good set of factor flows $q_{k,t}$ that explain common variation in asset flows. Empirically, this proposition implies that understanding the signs and magnitudes of flow betas is economically meaningful.

Orthogonalization. We orthogonalize the factor portfolios $b_{n,k}$ and factor flows $q_{n,k}$. The K orthogonalized factors have uncorrelated flows $\tilde{q}_{n,k}$ and the factor portfolios $\tilde{b}_{n,k}$ have uncorrelated fundamental risk. Appendix B.2 provides technical details.

Second-stage regression. We estimate the orthogonalized price impact model to obtain the price of risk $\tilde{\lambda}_k$. We denote the (n, m)-th element of fundamental-risk matrix $\text{var}(\mathbf{R}_0)$ as $v_{n,m}$. Under orthogonalized factors, model (14) implies the following panel regression

$$r_{n,t} = \sum_{k=1}^{K} \tilde{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right) + \xi_{n,t}. \tag{16}$$

The term $\sum_{m=1}^{N} v_{n,m}\tilde{b}_{m,k}$ is the covariance $\text{cov}(\mathbf{R}_0(n), \tilde{\mathbf{b}}_k^{\top}\mathbf{R}_0)$ between asset n's fundamental return and orthogonalized factor k's, so is the quantity of risk[17] per unit of flow $\tilde{q}_{k,t}$. The regressor $\tilde{q}_{k,t}\left(\sum_{m=1}^{N} v_{n,m}\tilde{b}_{m,k}\right)$ is the total quantity of risk induced by factor-k flow in period t. Because orthogonalized factors have uncorrelated flows and uncorrelated fundamental risks, each orthogonalized factor k has its own price of risk $\tilde{\lambda}_k$.

[17] Instead of pre-estimating fundamental-risk matrix $\text{var}(\mathbf{R}_0)$, we can equivalently form portfolios $\tilde{\mathbf{b}}_k^{\top}\mathbf{R}_0$ using the first-stage regression and orthogonalization. We then estimate each asset's fundamental-return beta to orthogonalized factors. The fundamental-return beta equals $\text{cov}(\mathbf{R}_0(n), \tilde{\mathbf{b}}_k^{\top}\mathbf{R}_0)$ because $\text{var}(\tilde{\mathbf{b}}_k^{\top}\mathbf{R}_0) = 1$.

Rotation to the original model. We recover the $K \quad K$ prices of risk $\{\lambda_{k,t}\}$ under the original model (14). Appendix B.2 provides technical details.

Standard Errors. Like the canonical Fama-MacBeth regression, our estimation also faces a generated-regressor issue. To compute standard errors, we use bootstrapping. We sample with replacement the panel $\{r_{n,t}, f_{n,t}, q_{k,t}\}$ across times $t = 1, 2, \ldots, T$. The bootstrapping is theoretically justified because our model imposes no intertemporal restrictions. For each bootstrapping sample, we compute the corresponding parameter estimates. We compute standard errors based on estimates from all bootstrapping samples.

3.4 Flow-Based Gibbons-Ross-Shanken Test

We test whether idiosyncratic flows create anomaly price impacts.

We run the second-stage regression (16) with additional idiosyncratic flows,

$$r_{n,t} = \sum_{m=1}^{N} a_{n,m} e_{m,t} + \sum_{k=1}^{K} \tilde{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right) + \xi_{n,t}. \tag{17}$$

The idiosyncratic flow $e_{m,t}$ of asset m at time t is the residual from the first-stage regression (15). The coefficient $a_{n,m}$ measures the anomaly price impact on asset n by the idiosyncratic flow of asset m. The true model implies the null hypothesis

$$H_0 : a_{n,m} = 0 \text{ for all } n \text{ and } m. \tag{18}$$

That is, the idiosyncratic flow of any asset m should not generate price impacts for any asset n, including itself. Let $\hat{\mathbf{a}}$ be the $N^2 \quad 1$ 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

$$T \hat{\mathbf{a}}^{>} \frac{\text{var}(\hat{\mathbf{a}})^{\ 1}}{T} \hat{\mathbf{a}} \quad \chi^2_{N^2}. \tag{19}$$

The cross section of price impacts. To understand the economic meaning of the χ^2

test statistic, we study the cross section of price impacts. Under our estimated model (16), the price impact of asset n at time t is $\delta_{n,t} = \sum_{k=1}^{K} \hat{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right)$, where we use $\hat{\lambda}_k$ to denote the estimates of $\tilde{\lambda}_k$. Following (8), the price impact ratio of any portfolio $\mathbf{c}$ *2* R^N at time t is $\mathbf{c}^\top \boldsymbol{\delta}_t / \sigma(\mathbf{c}^\top \mathbf{R}_0)$, where the cross section of price impact is $\boldsymbol{\delta}_t = (\delta_{1,t}, \delta_{2,t}, \ldots, \delta_{N,t})^\top$. The numerator is the price impact of portfolio $\mathbf{c}$ at time t and the denominator is the fundamental risk. The maximum price impact ratio (MPIR) across all portfolios is

$$\max_{\mathbf{c} 2 \mathsf{R}^N} \frac{\mathbf{c}^\top \boldsymbol{\delta}_t}{\sigma(\mathbf{c}^\top \mathbf{R}_0)} = \sqrt{\boldsymbol{\delta}_t^\top \operatorname{var}(\mathbf{R}_0)^{-1} \boldsymbol{\delta}_t}. \tag{20}$$

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

$$\theta^2 := \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\delta}_t^\top \operatorname{var}(\mathbf{R}_0)^{-1} \boldsymbol{\delta}_t. \tag{21}$$

Similarly, the price impact under the model (17) with anomaly price impacts[18] is $\check{\delta}_{n,t} = \sum_{m=1}^{N} \hat{a}_{n,m} e_{m,t} + \sum_{k=1}^{K} \hat{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right)$, where $\check{\boldsymbol{\delta}}_t = (\check{\delta}_{1,t}, \check{\delta}_{2,t}, \ldots, \check{\delta}_{N,t})^\top$. We define the time-series average of the realized squared MPIR as

$$\check{\theta}^2 := \frac{1}{T} \sum_{t=1}^{T} \check{\boldsymbol{\delta}}_t^\top \operatorname{var}(\mathbf{R}_0)^{-1} \check{\boldsymbol{\delta}}_t. \tag{22}$$

Appendix A.5 proves the following flow-based Gibbons, Ross, and Shanken (1989) result.

THEOREM 2. *Assuming that $e_{n,t}$, $\tilde{b}_{n,k}$, and $\tilde{q}_{k,t}$ are observed and the cross-sectional error term $\boldsymbol{\xi}_t$ is i.i.d. over time, we have, almost surely in the limit of T tending to in nity,*

$$\hat{\mathbf{a}}^\top \frac{\operatorname{var}(\hat{\mathbf{a}})^{-1}}{T} \hat{\mathbf{a}} = \check{\theta}^2 \quad \theta^2. \tag{23}$$

Equation (23) shows that, as period T tends to infinity, the χ^2 test statistic for the anomaly price impact in (19) equals the difference between the realized and model-implied

[18] Appendix A.4 proves that the estimated prices of risk $\hat{\lambda}_k$ remain unchanged in regressions (16) and (17).

average squared MPIR. This result is parallel to the original Gibbons, Ross, and Shanken (1989), who show that the χ^2 test statistic for the anomaly expected return equals the difference between the realized and model-implied squared maximum Sharpe ratio. Note that $e_{n,t}$, $\tilde{b}_{n,k}$, and $\tilde{q}_{k,t}$ are estimated from the first-stage regression and orthogonalization, so one should also account for the generated-regressor issue. A generalization of the Shanken (1992) correction is likely doable, and we leave this for future research.

The empirical implication of Theorem 2 is that, to *jointly* test the N^2 anomaly price impacts $a_{n,m}$ of all assets, which allows us to differentiate the price impacts of common flows from idiosyncratic flows, it suffices to study the corresponding MPIR portfolio.

4 Empirical Results: A Three-Factor Model of Price Impacts

4.1 Data and Empirical Measures

We measure flows using mutual fund flow-induced trading of Coval and Stafford (2007) and Lou (2012). We use the Fama-French 5 5 portfolios sorted on size and book-to-market equity as test assets, and measure their returns $r_{n,t}$ and flows $f_{n,t}$. We measure flows $q_{k,t}$ into the Fama-French three factors. We discuss flow construction in what follows.

First, we construct mutual fund flows in dollar amounts using standard procedure and also correct a number of data errors. Specifically, we obtain 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) for the sample period[19] of 2000 through September 2020. Our mutual fund sample contains 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 $r^{\text{fund}}_{m,t}$. We

[19]In 1990s, the mutual-fund industry experienced fast expansion and consistent inflows (Lou (2012); Li (2021); Ben-David, Li, Rossi, and Song (2021a)). After 2000, monthly mutual-fund flows are more stationary. Because we estimate a static model, we start the sample period in 2000, similar to Gabaix and Koijen (2022).

define the dollar-amount mutual fund flow as

$$f^{\text{fund}}_{m,t} = \text{TNA}_{m,t} \quad \text{TNA}_{m,t\ \ 1}(1 + r^{\text{fund}}_{m,t}). \tag{24}$$

We cross-check mutual funds' monthly returns and TNA from CRSP with those from Morningstar and Thomson/Refinitiv. We also manually correct a number of data input errors. Appendix C presents details. Correcting measurement errors is especially important because we measure flow in dollar amounts—a small error for a big fund causes significant bias.

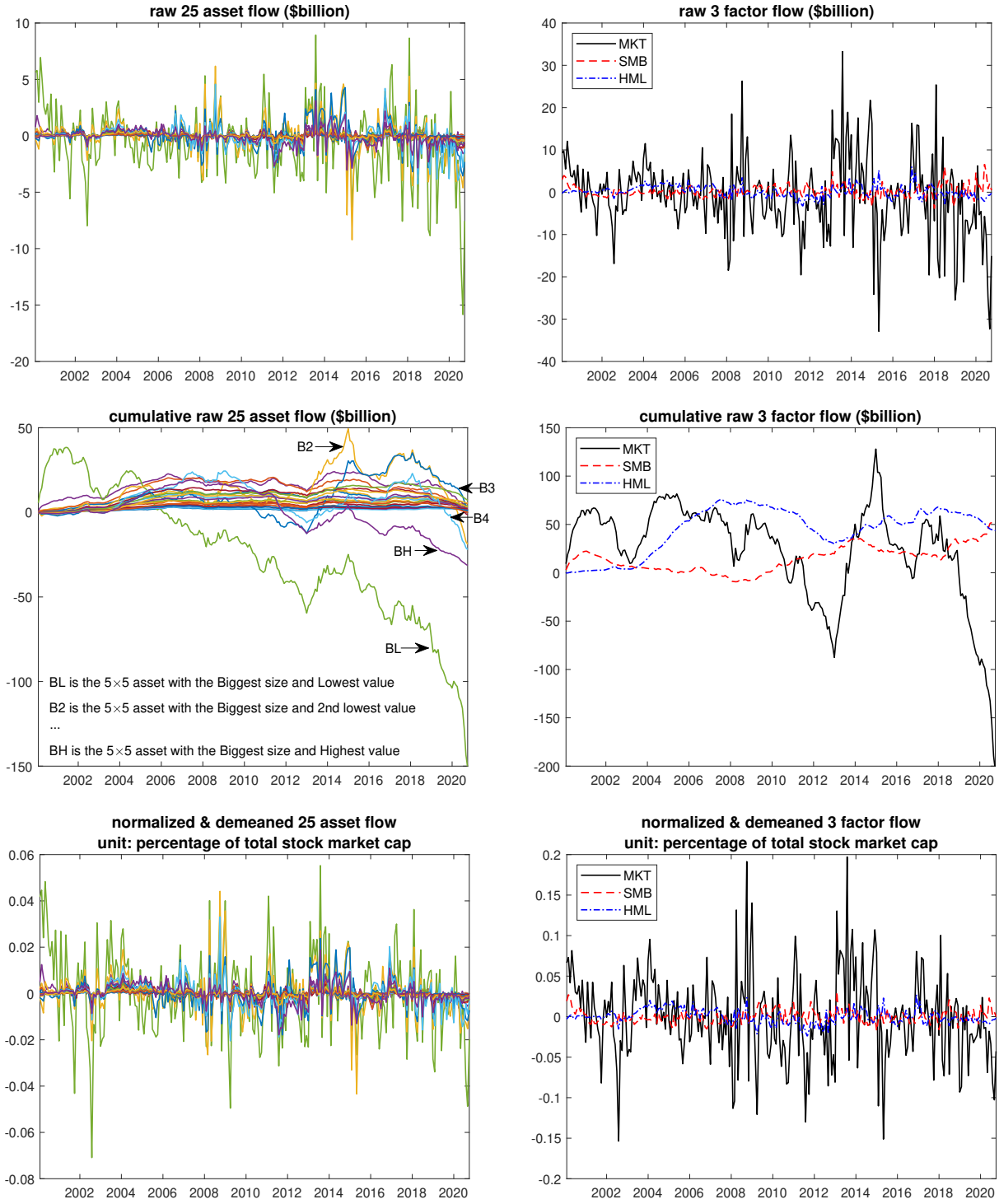
Second, we convert mutual fund flows to stock-level flows, using the standard assumption that mutual funds buy or sell stocks in proportion to their prior holdings. Importantly, we use the two-quarter-lagged mutual fund holding to convert mutual fund flow to stock-level flows. That is, for mutual fund flows in April, May, and June, we use the fund holding information as of Q4 the previous year. There are two reasons. First, the key feature of the flow-induced trading construction of Lou (2012) is lagged mutual fund holding. While mutual fund flows are mostly driven by uninformed retail investors, mutual fund managers may be informed. The lag ensures that the resulting stock-level flows proxy the non-discretionary component of mutual fund flows induced by retail investors. Second, we lag two quarters to ensure that the holding information is observable for our out-of-sample tradable strategy.[20]

Third, we use stock-level flows and equation (13) to construct the Fama-French MKT, SMB, and HML flows $q_{k,t}$ and 5 5 size and book-to-market equity sorted test asset flows $f_{n,t}$. The 25 test assets' monthly returns $r_{n,t}$ are obtained from Ken French's website.

In the top two panels of Figure 2, we plot the monthly test asset and factor flows. In the two middle panels, we plot the cumulative sum of test asset and factor flows. In terms of the 25 asset flows, big companies naturally exhibit the greatest fluctuation because of their large market capitalization. Asset flows exhibit strong commonality. The MKT factor exhibits the greatest fluctuation in flows, but SMB and HML flows also exhibit sizable variation.

[20]The holding information is reported with a maximum statutory delay of 45 days (Christoffersen, Danesh, and Musto (2015), which means that Q1 holding information is unlikely to be observable in April. We use the holding information as of Q4 the previous year for April, May, and June flows to be conservative.

Figure 2. Time series of 25 test asset flows and 3 factor flows



Note: In the top two figures, we plot the monthly test asset and factor flows. In the two middle figures, we plot the cumulative sum of test asset and factor flows. In the bottom two figures, we divide test asset and factor flows by the total stock market capitalization in the previous month and then subtract the unconditional time-series mean. The sample period runs from January 2000 through September 2020.

Finally, to account for the growing size of the total stock market, we normalize the asset and factor flows by total stock market capitalization in the previous month (defined as the sum of all 25 test assets). Differing from usual practices, we do not normalize asset or factor flows by their respective market capitalization. Our normalization is a consequence of our risk stance—price impacts arise because marginal investors absorb the *aggregate* payoff risk induced by flows. Our normalization preserves the relative dollar-amount magnitude between different flows in the cross section in order to correctly measure the aggregate risk. Using a simple example, the BL (Biggest-Lowest) test asset has about 40% of total stock market capitalization, and the SL (Smallest-Lowest) test asset has about 0.4%. For the BL asset, a 1% flow of its own market capitalization is 0.4% of total stock market capitalization; for the SL asset, a 1% flow is only 0.004%. From an aggregate risk perspective, the 1% BL asset flow is clearly much more important than the 1% SL asset flow, so we track these flows by 0.4% and 0.004% in the cross section. We also remove the unconditional time-series mean of each asset and factor flow, so we do not carry extra constants in the subsequent regressions.

The bottom two panels show the final test asset and factor flows. The MKT factor can experience inflow and outflow as large as 0.1% of total stock market capitalization.[21] SMB and HML exhibit smaller variations in flow. The correlation between the flows of (MKT, SMB), (MKT, HML), and (HML, SMB) are only 0.11, 0.25, and -0.11, respectively.

4.2 Common Variation in Flows

Table 1 reports the results of the first-stage regression (which is discussed in Section 3.3),

$$f_{n,t} = b_{n,\text{MKT}} q_{\text{MKT},t} + b_{n,\text{SMB}} q_{\text{SMB},t} + b_{n,\text{HML}} q_{\text{HML},t} + e_{n,t}. \tag{25}$$

[21] Flow into the MKT factor induced by mutual fund trading is significantly less volatile than aggregate mutual fund flow, which we show in Appendix Figure C.1. These are distinct objects. The variation in aggregate mutual fund flow consists of many factors, and flow into the MKT factor is just one of them.

This is a time-series regression of each asset n's flow $f_{n,t}$ on 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 ranges from 50% for small companies to 80% for big companies. Factor flows explain large portions of the common variation in asset flows.

Flow betas $b_{n,\text{MKT}}$, $b_{n,\text{SMB}}$, and $b_{n,\text{HML}}$ are *mimicking portfolio weights* for factor flows. Regression (25) implies that, holding all else equal, a one-dollar MKT-factor flow increases asset-n flow by $\$b_{n,\text{MKT}}$. Marginal investors are the rest of market who absorbs these flows. By market clearing, a one-dollar MKT-factor flow decreases marginal investors' holding in asset n by $\$b_{n,\text{MKT}}$. Therefore, the 25 flow betas $fb_{n,\text{MKT}}g$ mimic the *changes* in marginal investors' 5 5 portfolio holding caused by one dollar of MKT flow.

First, we study the flow beta of MKT factor $b_{n,\text{MKT}}$. That $b_{\text{BL,MKT}} = 0.3252$ means that if MKT flow increases by \$1, the BL asset flow increases by \$0.3252. The top-left panel of Table 1 reports the market capitalization weight w_n of the 5 5 assets, defined as the time-series average of asset n's market capitalization over the total stock market capitalization. The w_n proxies for the portfolio weight of the Fama-French MKT factor on the 5 5 test assets. For all assets n, flow beta $b_{n,\text{MKT}}$ is remarkably close to the Fama-French MKT weight w_n. All $b_{n,\text{MKT}}$ are positive, and their sum is very close to one. By Proposition 2, this empirical finding shows that asset flows indeed vary according to the MKT weights and the MKT flow $q_{\text{MKT},t}$ recovers this direction of variation.

Second, we study the flow beta of SMB factor $b_{n,\text{SMB}}$. Although the numbers not exactly the same, the flow betas $b_{n,\text{SMB}}$ share some similarity with the Fama-French small-minus-big construction. For example, for small companies, $b_{n,\text{SMB}}$ is positive; for big companies, $b_{n,\text{SMB}}$ is negative. Moreover, the sum of all positive $b_{n,\text{SMB}}$ is 0.99 and the sum of all negative $b_{n,\text{SMB}}$ is 0.89, so the absolute value of both sums are very close to one. Overall, the evidence supports that asset flows also have significant common variation in the SMB direction.

Third, we study the flow beta of HML factor $b_{n,\text{HML}}$. Only BL and B2 assets have negative flow betas, and all other assets have positive flow betas. The sum of positive $b_{n,\text{HML}}$ is almost

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 beta $b_{n,\text{MKT}}$					**SMB flow beta** $b_{n,\text{SMB}}$					**HML flow beta** $b_{n,\text{HML}}$				
	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 (25) 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 betas $b_{n,\text{MKT}}$, $b_{n,\text{SMB}}$, 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 and take the simple average over time. The t-statistics are calculated from heteroskedasticity-robust standard errors.

2 and the sum of negative $b_{n,\text{HML}}$ is about 0.5. Overall, the estimated flow betas have some Fama-French HML flavor, but the numbers differ. The evidence shows that the total quantity of flows in the HML direction is really BL and B2 assets against all other assets, because BL and B2 assets combined have over 50% of total stock market capitalization.

Overall, the first-stage evidence supports two results: (i) Fama-French factor flows are systematic flows and (ii) absorbing Fama-French factor flows exposes marginal investors to systematic risks. We next orthogonalize the Fama-French factor flows and risks.

4.3 Model Orthogonalization

We orthogonalize the MKT, SMB, and HML factors. The intuition is that the MKT, SMB, and HML factors could have cross impacts on each other because they have correlated flows and correlated fundamental risks. If we directly run a price impact regression, we would have 3 3 = 9 prices of risk for 3 factors. The orthogonalized MKT, SMB, and HML factors have uncorrelated flows and uncorrelated fundamental risks. The structural restriction, which is called Irrelevance of Uncorrelated Flows (IUF) and discussed in Section 3.1, implies that orthogonalized factors have no cross impacts and thus have in total 3 prices of risk.

Table 2 shows how we obtain the orthogonalized MKT, SMB, and HML factors $\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \tilde{\mathbf{b}}_3$ from the original MKT, SMB, and SMB factors. From the top-left panel, we see that MKT flow is about six times as volatile as SMB and HML flows, and their pairwise correlation is small but non-zero. In the top-right panel, we report the correlations and standard deviations of the one-year returns of mimicking portfolios, which are formed using the flow betas $\mathbf{b}_{\text{MKT}} = f b_{n,\text{MKT}} g$, $\mathbf{b}_{\text{SMB}} = f b_{n,\text{SMB}} g$, and $\mathbf{b}_{\text{HML}} = f b_{n,\text{HML}} g$. The one-year returns are used to proxy the fundamental risks of mimicking portfolios. We see that mimicking portfolios have highly correlated returns, especially between MKT and HML mimicking portfolios. This is because the estimated SMB and HML flow betas differ from the original Fama-French portfolio weights, as discussed in the preceding section.

The bottom panel shows the orthogonalized MKT, SMB, and HML factors $\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \tilde{\mathbf{b}}_3$,

Table 2. Model orthogonalization

original factors

	flow					return of mimicking portfolios			
	flow correlation			flow std.		return correlation			return std.
	MKT	SMB	HML			MKT	SMB	HML	
MKT	1	0.11	0.25	5.34	10^{-4}	1	0.60	0.87	0.16
SMB	0.11	1	-0.11	0.83	10^{-4}	0.60	1	0.67	0.11
HML	0.25	-0.11	1	0.87	10^{-4}	0.87	0.67	1	0.35

orthogonalized factors

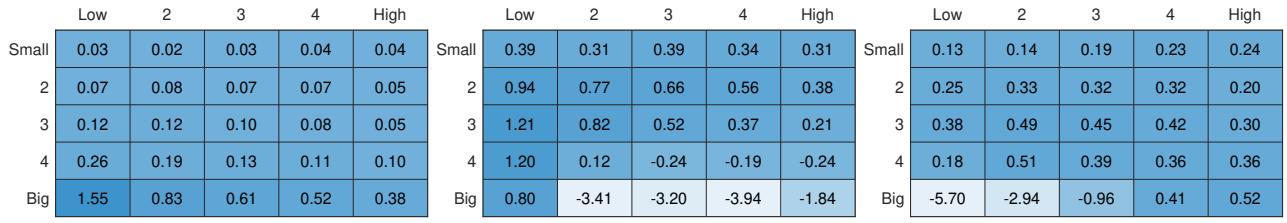
	flow					return of mimicking portfolios			
	flow correlation			flow std.		return correlation			return std.
	$\tilde{\mathbf{b}}_1$	$\tilde{\mathbf{b}}_2$	$\tilde{\mathbf{b}}_3$			$\tilde{\mathbf{b}}_1$	$\tilde{\mathbf{b}}_2$	$\tilde{\mathbf{b}}_3$	
$\tilde{\mathbf{b}}_1$	1	0	0	9.88	10^{-5}	1	0	0	1
$\tilde{\mathbf{b}}_2$	0	1	0	0.64	10^{-5}	0	1	0	1
$\tilde{\mathbf{b}}_3$	0	0	1	1.26	10^{-5}	0	0	1	1

rotation from original factors to orthogonalized factors

orthogonalized MKT: $\tilde{\mathbf{b}}_1 = 5.21 \quad \mathbf{b}_{\text{MKT}} + 0.10 \quad \mathbf{b}_{\text{SMB}} + 0.43 \quad \mathbf{b}_{\text{HML}}$

orthogonalized SMB: $\tilde{\mathbf{b}}_2 = \quad 0.59 \quad \mathbf{b}_{\text{MKT}} + 12.77 \quad \mathbf{b}_{\text{SMB}} \quad 2.39 \quad \mathbf{b}_{\text{HML}}$

orthogonalized HML: $\tilde{\mathbf{b}}_3 = \quad 11.39 \quad \mathbf{b}_{\text{MKT}} + 0.05 \quad \mathbf{b}_{\text{SMB}} + 5.90 \quad \mathbf{b}_{\text{HML}}$



$\tilde{\mathbf{b}}_1$ **(orthogonalized MKT)** $\tilde{\mathbf{b}}_2$ **(orthogonalized SMB)** $\tilde{\mathbf{b}}_3$ **(orthogonalized HML)**

Note: We report the correlations and standard derivations of flows and fundamental returns of mimicking portfolios for the original MKT, SMB, and HML factors and for the orthogonalized MKT, SMB, and HML factors $\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \tilde{\mathbf{b}}_3$. The unit of flow is the fraction of total stock market capitalization. We measure the fundamental risk using the one-year returns. We also report the rotation from the original factors $\mathbf{b}_{\text{MKT}}$, $\mathbf{b}_{\text{SMB}}$, and $\mathbf{b}_{\text{HML}}$ to the orthogonalized factors $\tilde{\mathbf{b}}_1$, $\tilde{\mathbf{b}}_2$, and $\tilde{\mathbf{b}}_3$. The three bottom figures report the portfolio weights of the orthogonalized factors $\tilde{\mathbf{b}}_1$, $\tilde{\mathbf{b}}_2$, and $\tilde{\mathbf{b}}_3$.

which have uncorrelated flows and uncorrelated fundamental risks. In particular, $\tilde{\mathbf{b}}_1 = 5.21 \quad \mathbf{b}_{\text{MKT}} + 0.10 \quad \mathbf{b}_{\text{SMB}} + 0.43 \quad \mathbf{b}_{\text{HML}}$ implies that $\tilde{\mathbf{b}}_1$ can be interpreted as the orthogonalized MKT factor. This interpretation is also evident by directly inspecting the portfolio weights $\tilde{\mathbf{b}}_1$ in the bottom-left figure. Similarly, $\tilde{\mathbf{b}}_2 = \quad 0.59 \quad \mathbf{b}_{\text{MKT}} + 12.77 \quad \mathbf{b}_{\text{SMB}} \quad 2.39 \quad \mathbf{b}_{\text{HML}}$

can be interpreted as the orthogonalized SMB factor. Lastly, $\tilde{\mathbf{b}}_3 = \quad 11.39 \quad \mathbf{b}_{\text{MKT}} + 0.05 \quad \mathbf{b}_{\text{SMB}} + 5.90 \quad \mathbf{b}_{\text{HML}}$ implies that $\tilde{\mathbf{b}}_3$ is the orthogonalized HML factor after removing the MKT component. From the portfolio weights $\tilde{\mathbf{b}}_3$ in the bottom-right figure, we see that the orthogonalized HML factor is about the BL, B2, and B3 test assets against all other assets. We next use the three orthogonalized factors to explain cross-sectional price impacts.

4.4 The Cross Section of Price Impacts

Table 3 reports the results of the second-stage regression (which is discussed in Section 3.3),

$$r_{n,t} = \sum_{k=1}^{3} \tilde{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right) + \xi_{n,t}. \tag{26}$$

Recall that $\tilde{q}_{k,t}$ is the month-t flow into the orthogonalized factor k. Additionally, we calculate the quantity of risk exposure $\text{cov}(\mathbf{R}_0(n), \tilde{\mathbf{b}}_k^{\top} \mathbf{R}_0)$ of each asset n to each orthogonalized factor $\tilde{\mathbf{b}}_k = \mathit{f}\tilde{b}_{n,k}\mathit{g}$. As mentioned before, we use one-year returns to measure the N assets' fundamental risk $\text{var}(\mathbf{R}_0)$ and we denote its (n,m)-th entry as $v_{n,m}$. Therefore, the quantity of risk is $\text{cov}(\mathbf{R}_0(n), \tilde{\mathbf{b}}_k^{\top} \mathbf{R}_0) = \sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k}$. We remove the unconditional time-series mean of return $r_{n,t}$ for each asset n, so regression (26) does not have constant terms. By regressing asset return $r_{n,t}$ on the product of factor flow $\tilde{q}_{k,t}$ and quantity of risk $\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k}$, this regression obtains each factor's price of risk $\tilde{\lambda}_k$.

The first column in panel A of Table 3 reports the estimated prices of risk $\tilde{\lambda}_k$. We report the heteroskedasticity robust t-statistics for the second-stage regression (26) and bootstrapped t-statistics. The bootstrapped t-statistics are significantly smaller, because $\tilde{q}_{k,t}$ and $\tilde{b}_{m,k}$ are generated regressors from the first-stage regression and orthogonalization.

The estimated $\tilde{\lambda}_1 = 673$ implies that if the flow into the first orthogonalized factor increases by one standard deviation ($9.98 \quad 10^{\ -5}$ of total stock market capitalization from Table 2), the factor's return increases by 6.6% ($= 673 \quad 9.98 \quad 10^{\ -5}$). Recall that the factor is normalized to have annualized return volatility 100%. Therefore, the flow-induced price

Table 3. Second-stage regression: returns on factor flows quantity of risk

	Panel A: price of risk of orthogonalized factors		
	total return OLS	day return OLS	day return IV
$\tilde{\lambda}_1$	673	429	508
$t(\tilde{\lambda}_1)$ (robust)	(14.41)	(12.02)	(3.03)
$t(\tilde{\lambda}_1)$ (bootstrap)	(3.20)	(2.78)	(0.78)
$\tilde{\lambda}_2$	19,211	9,119	37,757
$t(\tilde{\lambda}_2)$ (robust)	(6.44)	(3.57)	(3.04)
$t(\tilde{\lambda}_2)$ (bootstrap)	(2.61)	(1.52)	(1.26)
$\tilde{\lambda}_3$	3,309	2,447	3,297
$t(\tilde{\lambda}_3)$ (robust)	(5.54)	(4.56)	(1.14)
$t(\tilde{\lambda}_3)$ (bootstrap)	(1.55)	(1.34)	(0.29)
regression R^2	6.93%	3.86%	0.64%

Panel B: price of risk of original factors (rotating from total return OLS estimates in panel A)

	price of risk				bootstrapped t-stat		
	MKT	SMB	HML		MKT	SMB	HML
MKT	1,064	-1,647	-4,321	MKT	2.75	-0.59	-1.27
SMB	-293	19,069	-689	SMB	-0.94	2.59	-0.28
HML	-141	-3,028	3,060	HML	-0.88	-1.63	1.57

Note: In panel A, we run the second-stage regression of demeaned returns on factor flows times quantity of risk to obtain the price of risk of orthogonalized factors. The first two columns report the OLS estimation results using total returns and day (open-to-close) returns. The third column reports the IV estimation results using day returns. We instrument each factor flow by the factor's concurrent overnight return and the difference between one-month and half-year lagged flows. We report the heteroskedasticity robust t-statistics for the second-stage regression and bootstrapped t-statistics (10,000 times). In panel B, we report the price of risk of MKT, SMB, and HML factors by rotating from total return OLS estimates in panel A.

impacts account for 5.3% (= $(6.6\%)^2$ 12) of the first factor's total return variation. A similar calculation implies that flows account for 18.1% and 2.1% of the second and third factors' total return variation, respectively.[22] Despite the second factor has the least volatile flows out of all three factors, the factor's return is actually most subject to flows, because of the large price of risk $\tilde{\lambda}_2 = 19,211$. Recall that Table 2 shows that the second factor is the orthogonalized SMB factor. We conjecture that the high price of risk may be associated with the difficulty of trading against flows into small companies.

Asset-pricing intuitions imply that different factors can have different price of risk, mean-

[22]The calculations are $(19,211 \quad 0.64 \quad 10^{\ 5})^2 \quad 12 = 18.1\%$ and $(3,309 \quad 1.26 \quad 10^{\ 5})^2 \quad 12 = 2.1\%$.

ing that a one-dollar flow can differentially impact the price of different factors. However, the static CARA-Gaussian price impact model, which is the literature's workhorse model for empirical estimation and theory foundation of price impacts, imposes the restriction that all prices of risk must be the same in the cross section. The data reject this restriction. Our theory-founded model relaxes this unrealistic constraint of the static CARA-Gaussian model and grants greater empirical flexibility to modeling cross-sectional variation of price impacts.

Addressing endogeneity concern. Regression (26) requires that $\mathsf{E}[\tilde{q}_{k,t}\xi_{n,t}] = 0$ (note that the quantity of risk $\sum_{m=1}^{N} v_{n,m}\tilde{b}_{m,k}$ does not depend on t). Otherwise, the regression suffers endogeneity concerns and the estimated price of risk would capture components other than price impacts. Our construction of flows $\tilde{q}_{k,t}$ partly mitigates the endogeneity concern, because $\tilde{q}_{k,t}$ proxies the component of mutual fund flows that are due to mostly retail investors, not mutual fund managers' discretionary allocation (recall that we use the lagged fund holdings). Retail investors are less likely to be informed about $\xi_{n,t}$.

However, the endogenous concern may still arise if mutual fund investors chase concurrent returns, leading to over-estimation of the prices of risk. We care about the prices of risk because they later feed into our MPIR portfolio and trading strategy. If we have correctly estimated the prices of risk, the price impacts of the corresponding MPIR should revert 100%; if we have over-estimated the price of risk, the price impacts should revert less than 100%. The main way that we address endogeneity concern is to estimate the ratio of reversion to price impact in sample, use this ratio to shrink the estimated prices of risk, and use the shrunk prices of risk to construct the out-of-sample MPIR strategy.

That being said, we conduct several robustness checks to ensure that our main specification mostly captures the price impact components. First, we follow Li (2021) to substitute the asset return $r_{n,t}$ by the day return $r_{n,t}^{\text{fdayg}}$, which is the monthly aggregate of all open-to-close returns in each trading day.[23] The idea is that the day (open-to-close) returns are more

[23] We follow Lou, Polk, and Skouras (2019) to construct the open-to-close (day) and close-to-open (night) returns on each date t as $r_t^{\text{fdayg}} = \text{close}_t/\text{open}_t \quad 1$ and $r_t^{\text{fnightg}} = (1 + r_t^{\text{fclose-to-closeg}})/(1 + r_t^{\text{fdayg}}) \quad 1$. Daily close-to-close returns and open and close prices are obtained from CRSP. Daily close-to-close returns are adjusted for corporate actions such as stock splits.

likely to reflect institutions' trading and the night (close-to-open) returns are more retails' trading (Lou, Polk, and Skouras (2019, 2022); Bogousslavsky and Muravyev (2021)). To focus on the price impact driven by mutual fund flows, not retail investors' return chasing, we use day return $r_{n,t}^{\{\text{day}\}}$ instead of total return $r_{n,t}$ as the left-hand side of the second-stage regression (26). The second column in panel A of Table 3 reports the regression results with $r_{n,t}^{\{\text{day}\}}$. The estimated prices of risk are similar but smaller than those in the main specification, suggesting that the main specification may suffer some endogeneity concern.

Second, we supplement the day return OLS estimate with an IV strategy with two instruments. The first instrument for factor flow $\tilde{q}_{k,t}$ is the factor's concurrent night return $\sum_{m=1}^{25} \tilde{b}_{m,k} r_{m,t}^{\{\text{night}\}}$. The IV relevance condition is satisfied because factor flows are positively associated with concurrent night returns due to mutual-fund investors' return chasing. Moreover, because night and day returns are non-overlapping, the IV exclusion restriction also likely satisfies. (The non-overlapping requirement means that we can use only day return, not total return, as the regression left-hand side.) The second instrument is $\tilde{q}_{k,t\ \ 1} \quad \tilde{q}_{k,t\ \ 6}$, the difference between one-month and half-year lagged flows. The relevance arises from the serial correlation in $\tilde{q}_{k,t}$. The exclusion is because lagged flows are unlikely to be informative about the fundamental-return component of $\xi_{n,t}$. However, if we were to directly use $\tilde{q}_{k,t\ \ 1}$ as an instrument, the exclusion restriction could fail because price impacts generated by $\tilde{q}_{k,t\ \ 1}$ in month $t \quad 1$ would revert in month t, leading to $\mathsf{E}[\tilde{q}_{k,t\ \ 1}\xi_{n,t}] \neq 0$. Assuming that the price impacts of monthly flows linearly revert in half years (which we later show is a reasonable assumption), the difference between one-month and half-year lagged flows $\tilde{q}_{k,t\ \ 1} \quad \tilde{q}_{k,t\ \ 6}$ is orthogonal to the reversion component of $\xi_{n,t}$, so satisfies exclusion (see Appendix B.3 for detailed derivation). The third column reports the IV estimation results. The point estimates are similar to those reported in the first two columns, but the statistical power is low because of the generated-regressor issue.[24]

Overall, the two robustness checks support that our main OLS specification reported in

[24]In Appendix Table D.3, we report details of the first-stage IV regression results. The first-stage F-statistics are all around six, so the low statistical power could also be due to the potential weak instruments.

the first column mostly captures price impacts. We proceed with this main specification to construct the MPIR strategy. Appendix D.1 provides robustness checks using day returns.

Price of risk of MKT, SMB, and HML factors. Panel B of Table 3 reports the price of risk of MKT, SMB, and HML factors,

$$r_{n,t} = \sum_{k,l \in \{\text{MKT,SMB,HML}\}} \lambda_{k,l} q_{l,t} \left(\sum_{m=1}^{N} v_{n,m} b_{m,k} \right) + \xi_{n,t}. \tag{27}$$

We obtain $\lambda_{k,l}$ by rotating from the orthogonalized $\tilde{\lambda}_k$ of our main specification reported in the first column of panel A, not by running a regression.[25] We see that the off-diagonal terms are non-zero. For example, the -4,321 coefficient at the top-right corner implies that flow into the HML factor negatively affects the concurrent MKT return. That is, flow into growth relative to value stocks increases the MKT return. The MKT, SMB, and HML factors have correlated flows and fundamental returns, so their price impacts should not be viewed in factor-by-factor isolation when studying the cross-sectional variation of price impacts. In Appendix D.2, we use the factor-implied MPIR to decompose how MKT, SMB, and HML factors explain the cross section of price impacts. Consistent with the findings for orthogonalized factors, we find that the risk induced by SMB flows has the greatest explanatory power because of the high price of risk and the MKT and HML factors have roughly the same explanatory power.

4.5 Maximum-Price-Impact-Ratio Portfolio

We study the maximum-price-impact-ratio portfolio and its subsequent reversion.

First, we write down the model-implied MPIR portfolio weights. The estimated price

[25] In Appendix Table D.4, we provide a robustness check by directly running (27) as a regression with nine free parameters $\lambda_{k,l}$. We find similar point estimates, albeit with lower statistical significance.

impact of asset n in month t is the fitted value of the second-stage regression (26),

$$\delta_{n,t} := \sum_{k=1}^{3} \tilde{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right). \tag{28}$$

By Theorem 2, the model-implied MPIR portfolio (which maximizes the ratio of its price impact to its fundamental risk $\mathbf{c}^\top \boldsymbol{\delta}_t / \sigma(\mathbf{c}^\top \mathbf{R}_0)$, where $\boldsymbol{\delta}_t := (\delta_{1,t}, \delta_{2,t}, \ldots, \delta_{N,t})^\top$) is

$$\tilde{\mathbf{c}}_t = \sum_{k=1}^{3} \tilde{\lambda}_k \tilde{q}_{k,t} \tilde{\mathbf{b}}_k. \tag{29}$$

The MPIR portfolio $\tilde{\mathbf{c}}_t$ longs the orthogonalized factor portfolio $\tilde{\mathbf{b}}_k$ if the factor flow $\tilde{q}_{k,t}$ is positive and shorts otherwise. The dynamic MPIR portfolio changes every month t depending on the flows $\tilde{q}_{k,t}$. The magnitude of the long/short position is also proportional to the estimated price of risk $\tilde{\lambda}_k$.

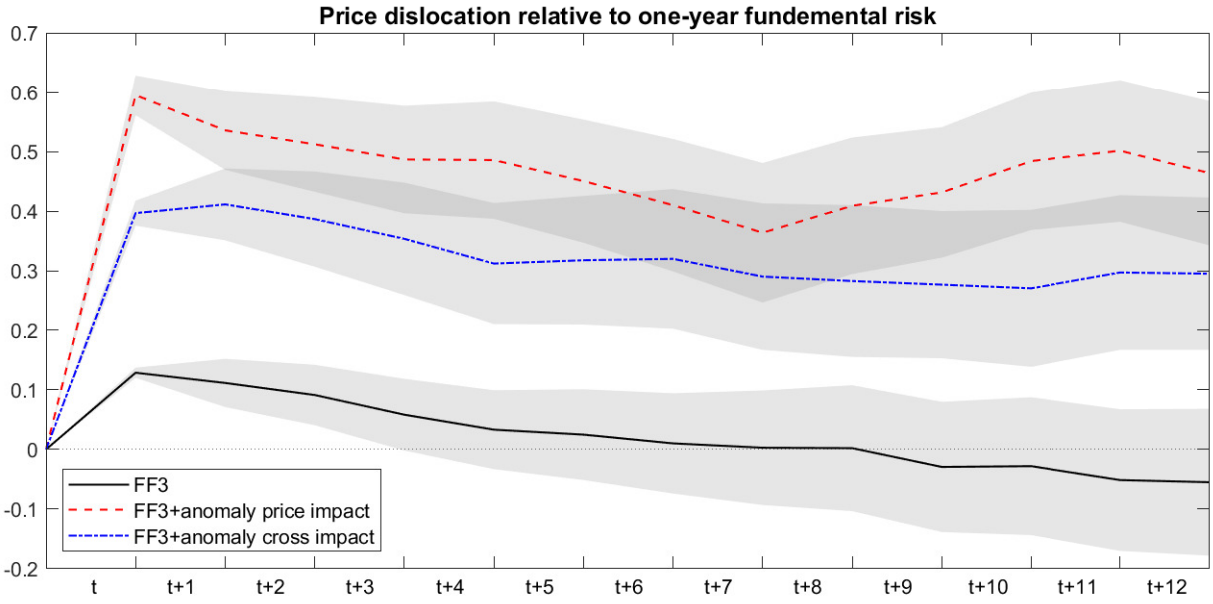
Next, we study the concurrent price impact and subsequent reversion of the MPIR portfolio. The price dislocation of the MPIR portfolio $\tilde{\mathbf{c}}_t$ between month t and $t+h$ (normalized by its fundamental risk) is defined as

$$\kappa_{t\,!\,t+h} = \frac{\overbrace{(\tilde{\mathbf{c}}_t)^\top \boldsymbol{\delta}_t}^{\text{price impact}} + \overbrace{(\tilde{\mathbf{c}}_t)^\top \mathbf{r}_{t+1\,!\,t+h}}^{\text{reversion}}}{\sigma((\tilde{\mathbf{c}}_t)^\top \mathbf{R}_0)}, \tag{30}$$

where $(\tilde{\mathbf{c}}_t)^\top \boldsymbol{\delta}_t$ is the concurrent month-t price impact of the MPIR portfolio and $(\tilde{\mathbf{c}}_t)^\top \mathbf{r}_{t+1\,!\,t+h}$ the subsequent month $t+1$ to $t+h$ price reversion. Specifically, $\mathbf{r}_{t+1\,!\,t+h}$ is the holding period return of the N assets between between month $t+1$ and $t+h$ (we remove the unconditional time-series mean of $\mathbf{r}_{t+1\,!\,t+h}$ to focus on the reversion component). The denominator is the annual return standard derivation of the MPIR portfolio.

The black solid line of Figure 3 plots the time-series means of $\kappa_{t\,!\,t+h}$ for each holding period $h = 0, 1, \ldots, 12$ on the horizontal axis, and the shaded area represents the 95% confidence intervals of the means. For $h = 0$, the MPIR portfolio's normalized price dislocation

Figure 3. Average price dislocation of the MPIR portfolio



Note: This figure shows the average price dislocation from month t to $t+h$ of the MPIR portfolio (normalized by its fundamental risk) for each holding period $h = 0, 1, \ldots, 12$ on the horizontal axis. The three lines correspond to three model specifications, our price impact model with MKT, SMB, and HML factor flows (FF3), FF3 plus anomaly price impact, and FF3 plus anomaly cross impact. The shaded area represents the 95% confidence intervals of the means. We measure fundamental risk using one-year returns. For each model, we compute the MPIR portfolio using equation (29). We then compute the normalized price dislocation of the MPIR portfolio between month t and $t + h$ using equation (30). Lastly, we compute the time-series means of the normalized price dislocation and the standard deviation of the mean for each horizon h.

is about 0.1 in an average month. That is, in the cross section, any portfolio's price impact is at most 10% of its annual return standard deviation, and the MPIR portfolio achieves this maximum. In about six subsequent months, the MPIR portfolio's price dislocation reverts to zero. This reversion is consistent with our initial stance that the price impacts of Fama-French factor flows are caused by marginal investors' risk aversion. In the next Section 4.6, we exploit the MPIR portfolio's price reversion to construct out-of-sample trading strategies.

Common versus idiosyncratic flows. Our factor model of price impacts (26) and its implied MPIR portfolio (29) depends only on factor flows $\tilde{q}_{k,t}$, but not on the idiosyncratic flows $e_{n,t}$ (the residual from the first-stage regression (25)). How should we understand the price impacts created by idiosyncratic flows? The flow-based Gibbons-Ross-Shanken Theorem 2 implies that we should compare the MPIR portfolio formed using factor-plus-

idiosyncratic flows with the MPIR portfolio that we have formed using only factor flows.

To form the factor-plus-idiosyncratic-flow MPIR portfolio, we run the second-stage regression (26) with additional idiosyncratic flows,

$$r_{n,t} = \check{\eta}_n e_{n,t} + \sum_{k=1}^{3} \tilde{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right) + \epsilon_{n,t}, \tag{31}$$

where the regression coefficient $\check{\eta}_n$ measures the anomaly impact of asset n's idiosyncratic flow $e_{n,t}$ on its own price. We present the results of this regression in Appendix Table D.2. Importantly, we use the regression fitted value $\check{\delta}_{n,t} := \check{\eta}_n e_{n,t} + \sum_{k=1}^{3} \tilde{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right)$ as the new price impact of asset n in month t and use this price impact to form the new MPIR portfolio.[26] We then apply the same equation (30) to compute the concurrent price impact and subsequent reversion of the MPIR portfolio.

The red dashed line in Figure 3 shows the time-series average price dislocation of factor-plus-idiosyncratic-flow MPIR portfolio. The concurrent price impact is significantly larger than that using only factor flows. The price dislocation of the factor-plus-idiosyncratic-flow MPIR portfolio barely reverts, however, even after one year. These two facts suggest that the larger concurrent price impact is likely caused by over-fitting, similar to the in-sample Markowitz optimization. (Note that our factor model (26) has 3 free parameters and the anomaly model (31) has $3 + 25$ parameters.) The lack of price reversion suggests that the anomaly price impacts caused by idiosyncratic flows are not driven by the risk channel. The sharp contrast between the black solid line and red dashed line shows that risk-driven price impacts depend only on factor flows but not on idiosyncratic flows. From a practical perspective, trading against idiosyncratic flows is unreliable, and forming a structured factor model is valuable.

As a further check, we also consider idiosyncratic flows' anomaly cross impacts. Specifically, we substitute asset n's idiosyncratic flow $e_{n,t}$ in regression (31) by the average idiosyn-

[26] Specifically, the new MPIR portfolio is $\check{\mathbf{c}}_t = \text{var}(\mathbf{R}_0)^{-1} \check{\boldsymbol{\delta}}_t$, where $\check{\boldsymbol{\delta}}_t = (\check{\delta}_{1,t}, \check{\delta}_{2,t}, \ldots, \check{\delta}_{N,t})^{\top}$.

cratic flow into asset n's neighboring 5 5 assets. The new regression measures the anomaly cross impacts created by neighboring assets' idiosyncratic flows (Appendix D.3 presents regression details). The blue dash-dotted line in Figure 3 again shows that the price dislocation of the new factor-plus-idiosyncratic-flow MPIR portfolio does not revert, suggesting that the anomaly cross impacts caused by idiosyncratic flows are also not driven by the risk channel.

4.6 Out-Of-Sample MPIR Strategy

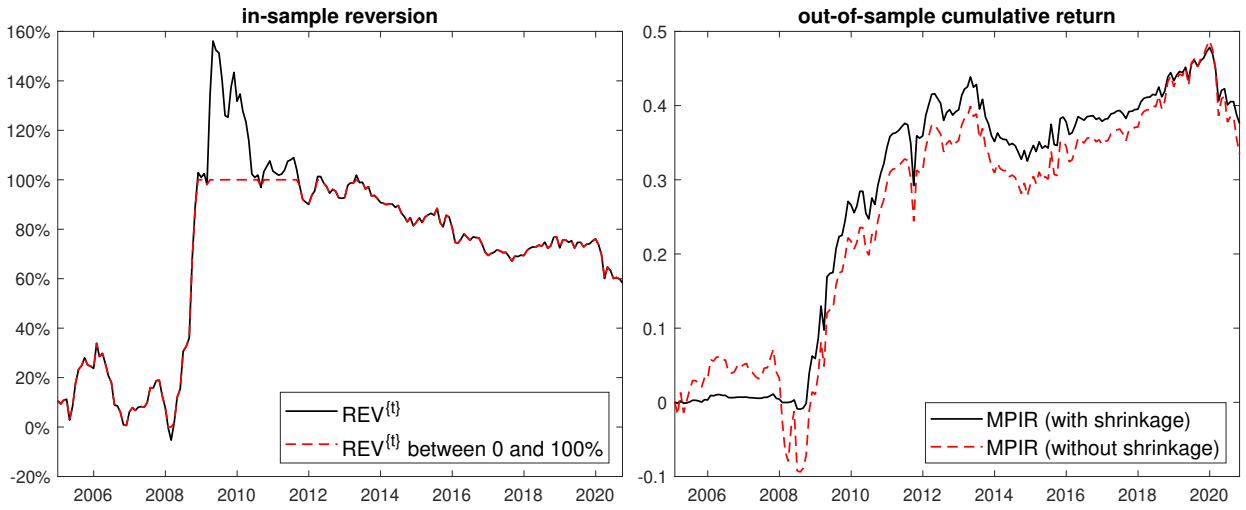
Having established the factor model of price impacts, we construct the model-implied optimal trading-against-MPIR strategy and examine its investment performance.

MPIR strategy construction. At the start of each month $t+1$, we estimate the factor model of price impacts using data up to month t. We form a staggered strategy that trades against the MPIR portfolios (constructed using equation (29)) of the past six months in equal weights. These MPIR portfolios incur the greatest price dislocation per unit of risk in months t 5 to t, and the staggered strategy is to take advantage of their subsequent price reversion. (We later consider the robustness of changing the starting and ending months of the staggered strategy.) Our MPIR strategy is effectively a monthly rebalancing portfolio on the Fama-French 5 5 test assets, by trading against the MKT, SMB, and HML flows into assets. The expanding estimation windows start from January 2000 to December 2004, and the out-of-sample testing period is from January 2005 to September 2020.

We first evaluate whether the MPIR strategy works in sample. For each training window, we run the strategy, compute the strategy's in-sample price reversion, and then compute the ratio of the average reversion to the average model-implied price impact. Ideally, this reversion ratio should be 100% according to theory (see Appendix B.4 for detailed formula).

The left panel of Figure 4 plots the ratio REV^{ftg} for each training window that ends in month t. For most of the windows, the ratio is less than 100%, meaning that the model-implied price impact does not fully revert. The reversion works significantly better since the 2008 financial crisis, consistent with the literature's finding that marginal investors are less

Figure 4. MPIR strategy's in-sample reversion and out-of-sample cumulative return



Note: In the left panel, we show REV^{ftg}, the ratio of the average one-month-forward reversion to the average model-implied price impact of the staggered MPIR portfolios for each training window that starts in January 2000 and ends in month t. In the right panel, we show the MPIR strategy's cumulative log returns with and without applying the shrinkage for the out-of-sample period from January 2005 to September 2020.

willing to absorb large quantity of risk post-crisis (Du, Tepper, and Verdelhan (2018)).

The fact that the in-sample reversion is less than 100% suggests that our model-implied price impact is too big. A likely reason is that the endogeneity concern discussed in Section 4.4 causes over-estimation of the prices of risk in the second-stage regression (26). To parsimoniously account for the over-estimation, for each training window, we shrink the estimated prices of risk by REV^{ftg}, if REV^{ftg} is smaller than 100%. If REV^{ftg} is negative, we assign zero prices of risk. These shrunk prices of risk, disciplined by the in-sample evaluation, are then used to construct the out-of-sample MPIR strategy for month $t+1$.

The right panel of Figure 4 plots the MPIR strategy's cumulative log returns. For the out-of-sample period from January 2005 to September 2020, the strategy's annualized Sharpe ratios are 0.48 and 0.34, with and without applying the shrinkage. We also compute the model-implied price impact ratio in each month. The strategy's annualized average price impact ratios are 0.52 and 0.77, with and without applying the shrinkage. Recall that the price impact ratio is a measure of the model-implied price dislocation, so if the price dislocation does fully revert, the average price impact ratio should be close to the out-of-

sample Sharpe ratio.[27] We see that the ratios are close only when applying the shrinkage, again showing that the shrinkage mitigates the over-estimation concern. Additionally, the MPIR strategy works better during the 2008 crisis, consistent with the idea that marginal investors require greater compensation for absorbing risk during the crisis, leading to better performance of reversion strategies (Nagel (2012)).

Combining MPIR strategy with fundamental investing. Although the MPIR strategy has a decent Sharpe ratio, we stress that the MPIR strategy is not designed to maximize the Sharpe ratio by itself. Rather, Theorem 1 shows that adding the MPIR strategy to fundamental investing improves the overall Sharpe ratio. We now empirically validate this central point.

We denote the MPIR strategy's excess return in month $t+1$ as $\tilde{r}_{t+1}$. (We later consider the robustness without applying the shrinkage.) We proxy fundamental investing using Jensen, Kelly, and Pedersen (2021) 154 anomaly portfolios, including Fama-French-Carhart six factors and a large list 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 combined excess return j in month $t+1$ is defined as

$$r_{j,t+1} = r_{j,t+1} + w_{j,t}\tilde{r}_{t+1}, \tag{32}$$

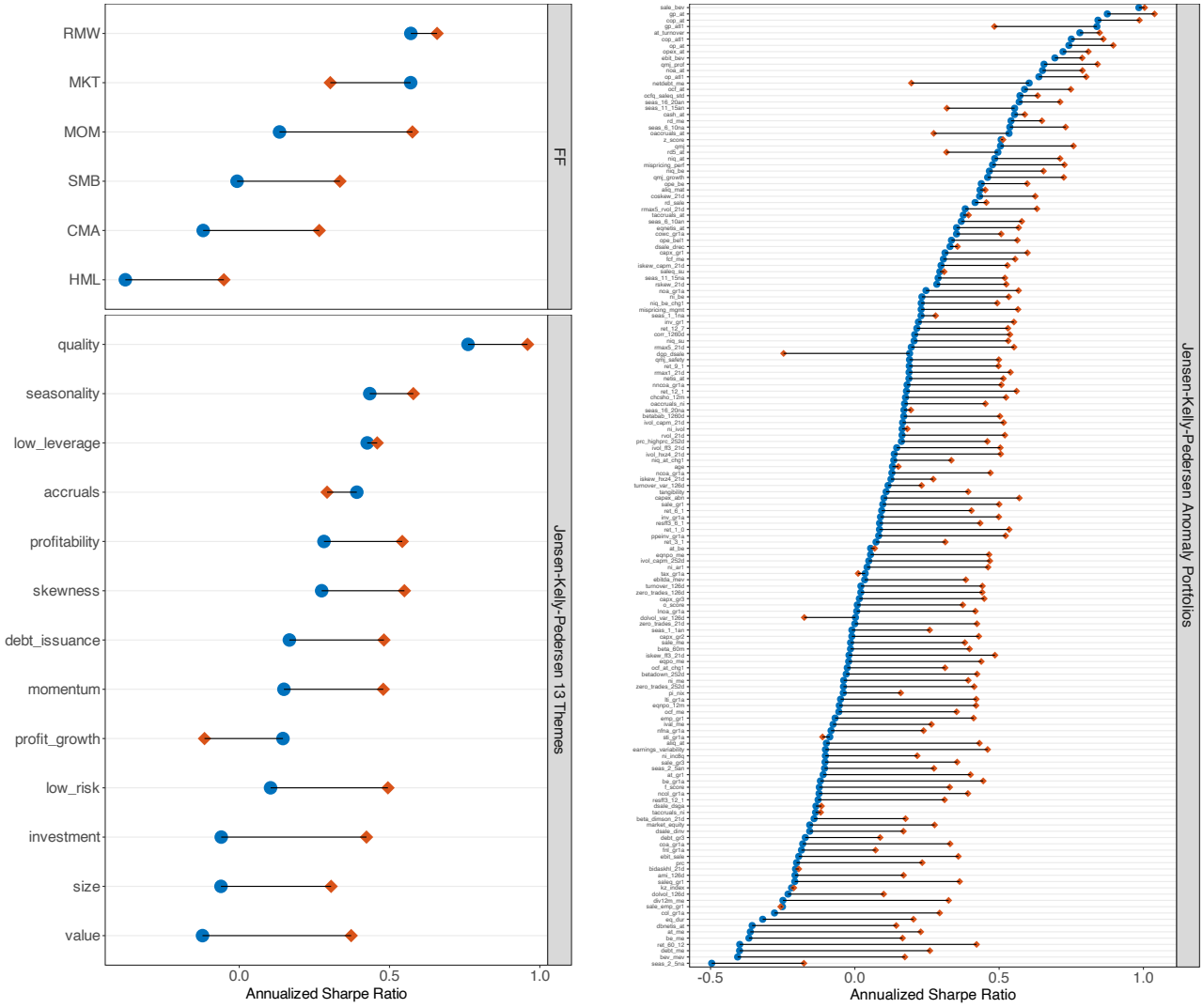
where $w_{j,t}$ is the mixing ratio estimated in sample according to Theorem 1, such that the return volatility of the anomaly portfolio j and the MPIR strategy are proportional to their respective Sharpe ratio and price impact ratio (see Appendix B.4 for detailed formula).

Figure 5 presents the Sharpe ratio improvement. The three panels show Fama-French-Carhart six factors, Jensen-Kelly-Pedersen 13 themes, and individual anomaly portfolios. The Sharpe ratios of the combined portfolios (red diamonds) are greater than those of the

[27] By Proposition 1, the MPIR strategy's Sharpe ratio equals the strategy's price impact ratio plus the Sharpe ratio driven by fundamental returns. Because the MPIR strategy is formed using flows, we expect the strategy's fundamental-return-driven Sharpe ratio to be close to zero on average.

[28] We use the 153 anomaly portfolios from Jensen, Kelly, and Pedersen (2021) and market excess return.

Figure 5. MPIR strategy increases out-of-sample Sharpe ratio of anomaly portfolios



	# of obs.	mean	std.	P5	P10	P25	median	P75	P90	P95
Sharpe ratio change	154	0.27	0.21	-0.08	0.02	0.15	0.32	0.42	0.51	0.56

Note: The blue dots represent the out-of-sample annualized Sharpe ratio of Jensen, Kelly, and Pedersen (2021) 154 portfolios, including Fama-French-Carhart six factors and a large list of other firm characteristics-based anomaly portfolios. The anomaly portfolios are also grouped by 13 themes. The red diamonds represent the Sharpe ratio of the portfolio that combines the original portfolio with our MPIR strategy. The MPIR strategy trades against the MPIR portfolios of the past six months in equal weights and applies the shrinkage for the estimated prices of risk. In each month, we scale the MPIR strategy such that the combined portfolio allocates risk between the original anomaly and our MPIR strategy proportional to their respective Sharpe ratio and price impact ratio. The expanding estimation windows start with January 2000 to December 2004, and the out-of-sample testing period is January 2005 to September 2020.

original portfolios (blue dots) across the board. Of the 154 portfolios, 143 or 93% receive a positive increase in Sharpe ratio. The average change in Sharpe ratio is 0.27 and the median change is 0.32. Mechanically speaking, the Sharpe ratio improvement arises from the low correlation between the MPIR strategy's return $\tilde{r}_{t+1}$ and the 154 anomaly portfolios' return $r_{j,t+1}$ (see Appendix Table D.5). The fact that the MPIR strategy consistently improves fundamental investing empirically validates the two-portfolio separation Theorem 1.

From a practical perspective, the MPIR strategy should be mostly immune to transaction cost and turnover concerns. First, the strategy by design trades against flows, or equivalently, provides liquidity to flows. By providing liquidity, the strategy is more likely to earn rather than pay the bid-ask spreads. Had we considered this effect, the performance evaluation could potentially be even better. Second, in terms of the turnover, our strategy trades against the staggered MPIR portfolios of past six months, so only one-sixth of the positions need to be changed every month.

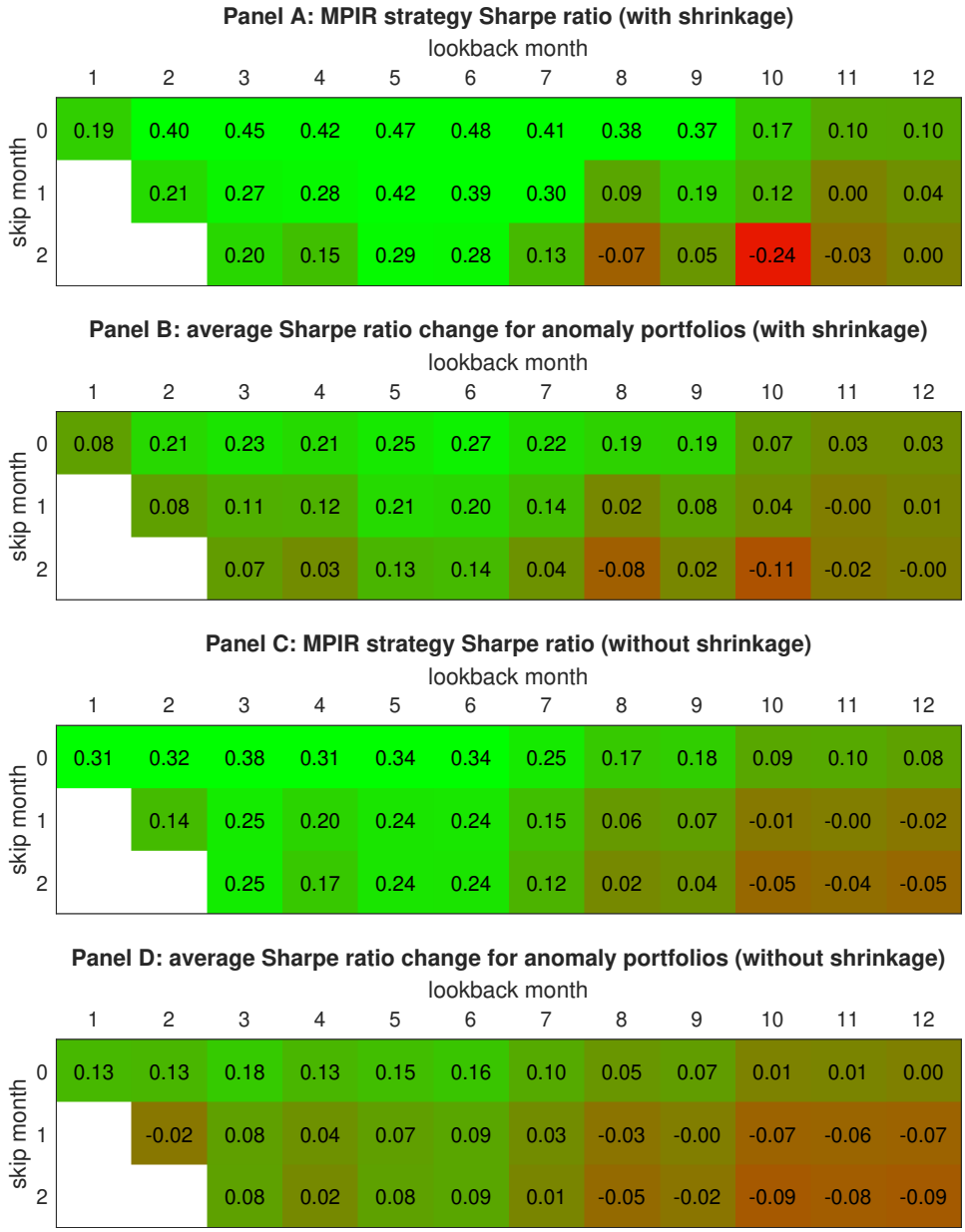
4.7 Robustness

We show that our MPIR strategy is robust to a series of alternative specifications.

Skip and lookback periods. In the main specification, our strategy trades against the average factor flows over the past six months. There might be concerns about whether monthly flow information is ready available at the start of next month. Additionally, whether we should look back six months or some other periods is unclear. We now show the results are robust to these concerns.

In the top two panels of Table 4, we report the Sharpe ratio of the MPIR strategy $\tilde{r}_{t+1}$ and the average Sharpe ratio change for the anomaly portfolios for different skip and lookback months. The investment performance is similar to our baseline specification, which corresponds to the (0,6) position of the table. In particular, even if we skip two months (that is, using up-to-January flow information at the end of March), our strategy still works decently. Unreported results for the median of Sharpe ratio changes are similar.

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



Note: In the top two panels, we report the MPIR strategy Sharpe ratio and the average Sharpe ratio change for the anomaly portfolios for different skip and lookback months. Our baseline specification that trades against the past six months' flows corresponds to the $(0, 6)$ position. The top two panels apply the shrinkage and the bottom two panels do not. The out-of-sample testing period is January 2005 to September 2020.

Applying the shrinkage or not. In the bottom two panels of Table 4, we replicate the top two panels but shut down the shrinkage. The MPIR strategy still works reasonably well. Nevertheless, the performance indeed slightly worsens without the shrinkage. This evidence is consistent with the idea that OLS estimates tend to over-estimate the price of risk and

the shrinkage is an effective approach to mitigate the over-estimation.

Using day returns to estimate price of risk. In Appendix D.1, we use day (open-to-close) returns as the left-hand side in the second-stage regression (26) to estimate the prices of risk and construct the MPIR strategy. We find similar investment performance.

5 Conclusion

We analyze the flow-driven fluctuations of the cross section of asset prices. We take the stance that price impacts of uninformative flows arise as marginal investors' risk compensation. We show that shorting the portfolio that incurs the maximum price impact for a given level of fundamental risk is the most efficient trading-against-flow strategy. To form this strategy, we build a new model of common factors of flows and common factors of fundamental returns and estimate the model using U.S. equity mutual fund flows into Fama-French three factors. Our strategy increases the out-of-sample annualized Sharpe ratio of 154 firm characteristics-based anomaly portfolios by an average of 0.3. Our evidence shows that risk-driven price impacts depend only on factor flows but not on idiosyncratic flows.

References

An, Yu, 2022, Flow-based arbitrage pricing theory, 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.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2021a, Ratings-driven demand and systematic price fluctuations, *Review of Financial Studies* Forthcoming.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2021b, What do mutual fund investors really care about? *Review of Financial Studies* Forthcoming.

Bogousslavsky, Vincent, and Dmitriy Muravyev, 2021, Who trades at the close? implications for price discovery and liquidity, Working paper, Boston College.

Bouchaud, Jean-Philippe, Julius Bonart, Jonathan Donier, and Martin Gould, 2018, *Trades, quotes and prices: nancial markets under the microscope* (Cambridge University Press).

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

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.

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).

Cochrane, John H, and Jesus Saa-Requejo, 2000, Beyond arbitrage: Good-deal asset price bounds in incomplete markets, *Journal of Political Economy* 108, 79–119.

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

Cremers, K. J. Martijn, and Jianping Mei, 2007, Turning over turnover, *Review of Financial Studies* 20, 1749–1782.

Da, Zhi, Borja Larrain, Clemens Sialm, and José Tessada, 2018, Destabilizing financial advice: Evidence from pension fund reallocations, *Review of Financial Studies* 31, 3720–3755.

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.

Dong, Xi, Namho Kang, and Joel Peress, 2022a, Fast and slow arbitrage: Fund flows and mispricing in the frequency domain, Working paper, City University of New York, Baruch College.

Dong, Xi, Karolina Krystyniak, and Lin Peng, 2022b, Liquidity shocks and institutional trading, Working paper, City University of New York, Baruch College.

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

Duffie, Darrell, 2012, Market making under the proposed Volcker rule, Working paper, Stanford University.

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

Frazzini, Andrea, Ronen Israel, and Tobias J Moskowitz, 2018, Trading costs, Working paper, Yale University.

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, Bryan Kelly, and Dacheng Xiu, 2021, Factor models, machine learning, and asset pricing, *Annual Review of Financial Economics* Forthcoming.

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

Greenwood, Robin, and David Thesmar, 2011, Stock price fragility, *Journal of Financial Economics* 102, 471–490.

Haddad, Valentin, Paul Huebner, and Erik Loualiche, 2022, How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing, Working paper, UCLA.

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hasbrouck, Joel, 1988, Trades, quotes, inventories, and information, *Journal of Financial Economics* 22, 229–252.

Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179–207.

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, 2021, Noise trading and asset pricing factors, Working paper, The University of Hong Kong.

Jensen, Theis Ingerslev, Bryan T Kelly, and Lasse Heje Pedersen, 2021, Is there a replication crisis in finance?, *Journal of Finance* Forthcoming.

Koijen, Ralph S. J., and Stijn van Nieuwerburgh, 2011, Predictability of returns and cash flows, *Annual Review of Financial Economics* 3, 467–491.

Koijen, Ralph SJ, Robert J Richmond, and Motohiro Yogo, 2021, Which investors matter for equity valuations and expected returns? Working paper, University of Chicago.

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, *The Journal of Finance* 73, 1183–1223.

Li, Jennifer Jie, Neil D Pearson, and Qi Zhang, 2020, Impact of demand shocks on the stock market: Evidence from Chinese IPOs, Working paper, INSEAD.

Li, Jiacui, 2021, What drives the size and value factors? Working paper, University of Utah.

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.

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.

Nagel, Stefan, 2021, *Machine learning in asset pricing* (Princeton University Press).

O'hara, Maureen, 1997, *Market microstructure theory* (Wiley).

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

Pavlova, Anna, and Taisiya Sikorskaya, 2022, Benchmarking intensity, Working paper, London Business School.

Peng, Cameron, and Chen Wang, 2021, Factor demand and factor returns, Working paper, LSE.

Ross, Stephen A, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–60.

Schmickler, Simon, 2020, Identifying the price impact of fire sales using high-frequency surprise mutual fund flows, Working paper, Princeton University.

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

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

Appendices

The appendices provide additional theoretical and empirical details.

A Proofs

We provide proofs omitted in the main text.

A.1 Proof of Proposition 1

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

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

A.2 Proof of Theorem 1

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

$$\mathrm{var}(\mathbf{R}(\mathbf{f}))^{\ 1} = \mathbf{W}(\mathbf{f})\mathrm{var}(\mathbf{R}(\mathbf{0}))^{\ 1}\mathbf{W}(\mathbf{f}). \qquad \text{(A.2)}$$

Second, we have

$$\mathbf{R}(\mathbf{0}) \quad R_F\mathbf{1} + R_F\mathbf{1} = \mathbf{W}(\mathbf{f})(\mathbf{R}(\mathbf{f}) \quad R_F\mathbf{1} + R_F\mathbf{1}), \qquad \text{(A.3)}$$

which simplifies to

$$\mathbf{R}(\mathbf{f}) \quad R_F\mathbf{1} = \mathbf{W}(\mathbf{f})^{\ 1}(\mathbf{R}(\mathbf{0}) \quad R_F\mathbf{1} \quad R_F\Delta\mathbf{p}(\mathbf{f})). \qquad \text{(A.4)}$$

Taking expectation on both sides, we have

$$\mathsf{E}[\mathbf{R}(\mathbf{f})] \quad R_F\mathbf{1} = \mathbf{W}(\mathbf{f})^{\ 1}(\mathsf{E}[\mathbf{R}(\mathbf{0})] \quad R_F\mathbf{1} \quad R_F\Delta\mathbf{p}(\mathbf{f})). \qquad \text{(A.5)}$$

Therefore, by equations (A.2) and (A.5), we have

$$\begin{aligned}\mathbf{c}\ (\mathbf{f}) &= \mathbf{W}(\mathbf{f})^{\ 1}\mathrm{var}(\mathbf{R}(\mathbf{f}))^{\ 1}(\mathsf{E}[\mathbf{R}(\mathbf{f})] \quad R_F\mathbf{1}) \\ &= \mathrm{var}(\mathbf{R}(\mathbf{0}))^{\ 1}(\mathsf{E}[\mathbf{R}(\mathbf{0})] \quad R_F\mathbf{1} \quad R_F\Delta\mathbf{p}(\mathbf{f})) = \mathbf{c}\ (\mathbf{0}) \quad R_F\tilde{\mathbf{c}}\ (\mathbf{f}). \qquad \text{(A.6)}\end{aligned}$$

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

$$\sigma(\mathbf{c}\ (\mathbf{0})^{>}\mathbf{R}(\mathbf{0})) = \sqrt{\mathbf{c}\ (\mathbf{0})^{>}\text{var}(\mathbf{R}(\mathbf{0}))\mathbf{c}\ (\mathbf{0})} = \sqrt{\mathsf{E}[\mathbf{R}(\mathbf{0}) \quad R_F\mathbf{1}]^{>}\text{var}(\mathbf{R}(\mathbf{0}))\ ^{1}\mathsf{E}[\mathbf{R}(\mathbf{0}) \quad R_F\mathbf{1}]}, \tag{A.7}$$

which equals the maximum Sharpe ratio without flow by definition (9). Similarly, the return volatility of portfolio $\tilde{\mathbf{c}}\ (\mathbf{f})$ is

$$\sigma(\tilde{\mathbf{c}}\ (\mathbf{f})^{>}\mathbf{R}(\mathbf{0})) = \sqrt{\tilde{\mathbf{c}}\ (\mathbf{f})^{>}\text{var}(\mathbf{R}(\mathbf{0}))\tilde{\mathbf{c}}\ (\mathbf{f})} = \sqrt{\Delta\mathbf{p}(\mathbf{f})^{>}\text{var}(\mathbf{R}(\mathbf{0}))\ ^{1}\Delta\mathbf{p}(\mathbf{f})}, \tag{A.8}$$

which equals the maximum price impact ratio by definition (8).

A.3 Proof of Proposition 2

We write the asset flows as an $N \quad T$ matrix $\mathbf{F} = \mathsf{f}f_{n,t}\mathsf{g}$ and the factor flows as an $K \quad T$ matrix $\mathbf{Q} = \mathsf{f}q_{k,t}\mathsf{g}$. We start by simplifying $\mathbf{B} = \mathbf{W}$. Equation (13) implies that

$$\mathbf{Q} = (\mathbf{W}^{>}\mathbf{W})\ ^{1}\mathbf{W}^{>}\mathbf{F}. \tag{A.9}$$

The first-stage time-series regression (15) then implies that

$$\mathbf{B}^{>} = (\mathbf{Q}\mathbf{Q}^{>})\ ^{1}\mathbf{Q}\mathbf{F}^{>}. \tag{A.10}$$

Combining the two equations above, we have

$$\mathbf{B} = \mathbf{F}\mathbf{F}^{>}\mathbf{W}(\mathbf{W}^{>}\mathbf{F}\mathbf{F}^{>}\mathbf{W})\ ^{1}\mathbf{W}^{>}\mathbf{W}. \tag{A.11}$$

We denote $\mathbf{S} = \mathbf{F}\mathbf{F}^{>}$, which is $N \quad 1$ times the sample covariance matrix of $f_{n,t}$. Therefore, $\mathbf{B} = \mathbf{W}$ is equivalent to

$$\mathbf{W} = \mathbf{S}\mathbf{W}(\mathbf{W}^{>}\mathbf{S}\mathbf{W})\ ^{1}\mathbf{W}^{>}\mathbf{W}. \tag{A.12}$$

First, we show that if $\mathbf{W}$ satisfies equation (A.12), then $\mathbf{W}$ can be rotated into a set of eigenvectors of $\mathbf{S}$. To see this, we conduct Cholesky decomposition and obtain $\mathbf{W}^{>}\mathbf{W} = \mathbf{U}^{>}\mathbf{U}$, where $\mathbf{U}$ is a $K \quad K$ upper triangular matrix. Equation (A.12) then simplifies to

$$\mathbf{W}\mathbf{U}\ ^{1} = \mathbf{S}\mathbf{W}\mathbf{U}\ ^{1}\mathbf{U}(\mathbf{W}^{>}\mathbf{S}\mathbf{W})\ ^{1}\mathbf{U}^{>}. \tag{A.13}$$

We define $\mathbf{Z} = \mathbf{W}\mathbf{U}^{-1}$ and obtain

$$\mathbf{Z} = \mathbf{S}\mathbf{Z}(\mathbf{Z}^{\top}\mathbf{S}\mathbf{Z})^{-1}. \tag{A.14}$$

We conduct eigenvalue decomposition and obtain

$$\mathbf{Z}^{\top}\mathbf{S}\mathbf{Z} = \mathbf{G}\mathbf{\Lambda}\mathbf{G}^{\top}, \tag{A.15}$$

where $\mathbf{G}$ is a $K \times K$ orthonormal matrix and $\mathbf{\Lambda}$ is a $K \times K$ diagonal matrix. Therefore,

$$(\mathbf{Z}\mathbf{G})\mathbf{\Lambda} = \mathbf{S}(\mathbf{Z}\mathbf{G}). \tag{A.16}$$

In other words, $\mathbf{Z}\mathbf{G} = \mathbf{W}\mathbf{U}^{-1}\mathbf{G}$ is K eigenvectors of $\mathbf{S}$.

Conversely, we show that if $\mathbf{W}$ can be rotated into a set of eigenvectors of $\mathbf{S}$, then $\mathbf{W}$ satisfies equation (A.12). To see this, we write $\mathbf{W} = \mathbf{X}\mathbf{O}$, where $\mathbf{O}$ is some $K \times K$ invertible matrix and $\mathbf{X}$ is K eigenvectors of $\mathbf{S}$ satisfying

$$\mathbf{X}\mathbf{\Lambda} = \mathbf{S}\mathbf{X} \text{ and } \mathbf{X}^{\top}\mathbf{X} = \mathbf{I}_k. \tag{A.17}$$

Rearranging the equation above, we obtain

$$\mathbf{X} = \mathbf{S}\mathbf{X}(\mathbf{X}^{\top}\mathbf{S}\mathbf{X})^{-1}. \tag{A.18}$$

Because $\mathbf{O}$ is invertible, we can further rearrange this equation into

$$\mathbf{X}\mathbf{O} = \mathbf{S}\mathbf{X}\mathbf{O}(\mathbf{O}^{\top}\mathbf{X}^{\top}\mathbf{S}\mathbf{X}\mathbf{O})^{-1}\mathbf{O}^{\top}\mathbf{X}^{\top}\mathbf{X}\mathbf{O}. \tag{A.19}$$

Note that $\mathbf{W} = \mathbf{X}\mathbf{O}$ and equation (A.19) is precisely equation (A.12).

Lastly, we prove that $\mathbf{W}$ can be rotated into a set of eigenvectors of $\mathbf{S}$ if and only if the resulting $\mathbf{Q}$ from (A.9) can be rotated into K principal components of asset flows $f_{n,t}$.

To see the only-if direction, we write $\mathbf{W} = \mathbf{X}\mathbf{O}$, where $\mathbf{O}$ is some $K \times K$ invertible matrix and $\mathbf{X}$ is K eigenvectors of $\mathbf{S}$. Note that

$$\mathbf{Q} = (\mathbf{W}^{\top}\mathbf{W})^{-1}\mathbf{W}^{\top}\mathbf{F} = (\mathbf{O}^{\top}\mathbf{X}^{\top}\mathbf{X}\mathbf{O})^{-1}\mathbf{O}^{\top}\mathbf{X}^{\top}\mathbf{F} = \mathbf{O}^{-1}\mathbf{X}^{\top}\mathbf{F} = \mathbf{O}^{-1}\tilde{\mathbf{Q}}, \tag{A.20}$$

where $\tilde{\mathbf{Q}} = \mathbf{X}^{\top}\mathbf{F}$ is precisely K principal components of asset flows $f_{n,t}$.

To see the if direction, we write $\mathbf{Q} = \mathbf{O}^{-1}\mathbf{X}^{\top}\mathbf{F}$, where $\mathbf{O}$ is some $K \times K$ invertible matrix

and $\mathbf{X}$ is K eigenvectors of $\mathbf{S}$. By equation (A.9), we have

$$(\mathbf{W}^{\top}\mathbf{W})\ {}^{1}\mathbf{W}^{\top} = \mathbf{O}\ {}^{1}\mathbf{X}^{\top}. \tag{A.21}$$

We define $\tilde{\mathbf{X}} = \mathbf{W}\mathbf{O}\ {}^{1}$, substitute into the equation above, and obtain

$$(\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}})\ {}^{1}\tilde{\mathbf{X}}^{\top} = \mathbf{X}^{\top}. \tag{A.22}$$

Because $\mathbf{X}$ are eigenvectors, we have

$$\mathbf{I}_K = \mathbf{X}^{\top}\mathbf{X} = (\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}})\ {}^{1}. \tag{A.23}$$

Therefore, we know that $\tilde{\mathbf{X}} = \mathbf{X}$, which implies that $\mathbf{W} = \mathbf{X}\mathbf{O}$.

A.4 Proof of the Simplifying Regression

We show that the panel regression (17),

$$r_{n,t} = \sum_{m=1}^{N} a_{n,m} e_{m,t} + \sum_{k=1}^{K} \tilde{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right) + \xi_{n,t}, \tag{A.24}$$

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

$$r_{n,t} = \sum_{m=1}^{N} a_{n,m} e_{m,t} + \xi_{n,t}. \tag{A.25}$$

obtains the same $a_{n,m}$ as regression (A.24). The second panel regression

$$r_{n,t} = \sum_{k=1}^{K} \tilde{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right) + \xi_{n,t}, \tag{A.26}$$

obtains the same $\tilde{\lambda}_k$ as regression (A.24).

To see this fact, note that because the idiosyncratic flow $e_{m,t}$ is the residual of the first-stage regression (15), 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 then stack the panel regression (A.24) in vector form and define the $NT \quad 1$ vector

$$\mathbf{r} = (r_{1,1}, r_{1,2}, \ldots, r_{1,T}, r_{2,1}, r_{2,2}, \ldots, r_{2,T}, \ldots, r_{N,1}, r_{N,2}, \ldots, r_{N,T})^{\top}. \tag{A.27}$$

We can rewrite the panel regression (A.24) in vector form as

$$\mathbf{r} = \sum_{n=1}^{N}\sum_{m=1}^{N} a_{n,m}\mathbf{e}_{n,m} + \sum_{k=1}^{K} \tilde{\lambda}_k \mathbf{y}_k + \boldsymbol{\xi}. \tag{A.28}$$

Each vector $\mathbf{e}_{n,m}$ is an $NT \quad 1$ vector with only the $(n \quad 1)T + 1$-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 $NT \quad 1$ vector

$$\mathbf{y}_k = \left(\underbrace{\tilde{q}_{k,1}\sum_{m=1}^{N} v_{1,m}\tilde{b}_{m,k}, \ldots, \tilde{q}_{k,T}\sum_{m=1}^{N} v_{1,m}\tilde{b}_{m,k}}_{T \text{ terms}}, \ldots, \underbrace{\tilde{q}_{k,1}\sum_{m=1}^{N} v_{N,m}\tilde{b}_{m,k}, \ldots, \tilde{q}_{k,T}\sum_{m=1}^{N} v_{N,m}\tilde{b}_{m,k}}_{T \text{ terms}} \right)^{\top}. \tag{A.29}$$

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

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

$$\mathbf{e}_{n,m}^{\top}\mathbf{y}_k = \sum_{l=1}^{N} v_{n,l}\tilde{b}_{l,k}\sum_{t=1}^{T}\tilde{q}_{k,t}e_{m,t} = 0, \tag{A.30}$$

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

$$\mathbf{r} = \sum_{n=1}^{N}\sum_{m=1}^{N} a_{n,m}\mathbf{e}_{n,m} + \boldsymbol{\xi} \text{ and } \mathbf{r} = \sum_{k=1}^{K} \tilde{\lambda}_k \mathbf{y}_k + \boldsymbol{\xi}. \tag{A.31}$$

This first regression further reduces to the asset-by-asset time-series regression (A.25).

A.5 Proof of Theorem 2

First, we simplify the χ^2 test statistics in (23). We write a $1 \quad N$ vector $\mathbf{e}_t = (e_{1,t}, e_{2,t}, \ldots, e_{N,t})$ and an $N \quad 1$ vector $\mathbf{a}_n = (a_{n,1}, a_{n,2}, \ldots, a_{n,N})^{\top}$, and we write regression (17) as

$$r_{n,t} = \mathbf{e}_t\mathbf{a}_n + \sum_{k=1}^{K}\tilde{\lambda}_k\left(\sum_{m=1}^{N} v_{n,m}\tilde{b}_{m,k}\right)\tilde{q}_{k,t} + \xi_{n,t}. \tag{A.32}$$

We define the $T \quad N$ matrix $\mathbf{e} = (\mathbf{e}_1; \mathbf{e}_2; \ldots; \mathbf{e}_T)$ and the $N \quad 1$ vector $\boldsymbol{\xi}_n = (\xi_{n,1}, \xi_{n,2}, \ldots, \xi_{n,T})^{\top}$. We show in Appendix A.4 that we can run an asset-by-asset time-series regression to obtain

the point estimator of $\mathbf{a}_n$ as

$$\hat{\mathbf{a}}_n = (\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\begin{pmatrix} r_{n,1} \\ r_{n,2} \\ \cdots \\ r_{n,T} \end{pmatrix} = \mathbf{a}_n + (\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\boldsymbol{\xi}_n, \tag{A.33}$$

where we have used 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,

$$\text{cov}(\hat{\mathbf{a}}_n, \hat{\mathbf{a}}_m) = (\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\text{cov}(\boldsymbol{\xi}_n, \boldsymbol{\xi}_m)\mathbf{e}(\mathbf{e}^{\top}\mathbf{e})^{-1} = (\mathbf{e}^{\top}\mathbf{e})^{-1}\boldsymbol{\Sigma}_{\xi}(n, m), \tag{A.34}$$

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 variance-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 $N^2 \quad 1$ vector of parameter estimates. By equation (A.34), we have

$$\text{var}(\hat{\mathbf{a}}) = \hat{\boldsymbol{\Sigma}}_{\xi} \quad (\mathbf{e}^{\top}\mathbf{e})^{-1}, \tag{A.35}$$

where represents the Kronecker product. Therefore, we have

$$\hat{\mathbf{a}}^{\top}\frac{\text{var}(\hat{\mathbf{a}})^{-1}}{T}\hat{\mathbf{a}} = \hat{\mathbf{a}}^{\top}\left(\hat{\boldsymbol{\Sigma}}_{\xi}^{-1} \quad (\mathbf{e}^{\top}\mathbf{e}/T)\right)\hat{\mathbf{a}}. \tag{A.36}$$

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

$$\hat{\mathbf{a}}^{\top}\left(\hat{\boldsymbol{\Sigma}}_{\xi}^{-1} \quad (\mathbf{e}^{\top}\mathbf{e}/T)\right)\hat{\mathbf{a}} = \sum_{n=1}^{N}\sum_{m=1}^{N}\left((\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\boldsymbol{\xi}_n\right)^{\top}\hat{\boldsymbol{\Sigma}}_{\xi}^{-1}(n, m)(\mathbf{e}^{\top}\mathbf{e}/T)(\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\boldsymbol{\xi}_m \tag{A.37}$$

$$= \frac{1}{T}\sum_{n=1}^{N}\sum_{m=1}^{N}\hat{\boldsymbol{\Sigma}}_{\xi}^{-1}(n, m)\boldsymbol{\xi}_n^{\top}\mathbf{e}(\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\boldsymbol{\xi}_m$$

$$= \frac{1}{T}\sum_{n=1}^{N}\sum_{m=1}^{N}\hat{\boldsymbol{\Sigma}}_{\xi}^{-1}(n, m)\boldsymbol{\psi}_n^{\top}\boldsymbol{\psi}_m \tag{A.38}$$

$$= \frac{1}{T}\begin{pmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 \\ \cdots \\ \boldsymbol{\psi}_N \end{pmatrix}^{\top}\left(\hat{\boldsymbol{\Sigma}}_{\xi}^{-1} \quad \mathbf{I}_T\right)\begin{pmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 \\ \cdots \\ \boldsymbol{\psi}_N \end{pmatrix}, \tag{A.39}$$

where we define

$$\boldsymbol{\psi}_n = \mathbf{e}(\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\boldsymbol{\xi}_n, \tag{A.40}$$

as the projection of $\boldsymbol{\xi}_n$ onto the idiosyncratic flow space. In step (A.37), we use block matrix multiplication for every N elements and $\hat{\boldsymbol{\Sigma}}_{\xi}^{-1}(n,m)$ is the (n,m)-th element of $\hat{\boldsymbol{\Sigma}}_{\xi}^{-1}$. In step (A.38), we use the fact that the projection matrix $\mathbf{e}(\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}$ is idempotent. In step (A.39), we use the block-matrix multiplication in 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

$$\frac{1}{T}\begin{pmatrix}\boldsymbol{\psi}_1\\ \boldsymbol{\psi}_2\\ \ldots\\ \boldsymbol{\psi}_N\end{pmatrix}^{\top}\begin{pmatrix}\hat{\boldsymbol{\Sigma}}_{\xi}^{-1} & \mathbf{I}_T\end{pmatrix}\begin{pmatrix}\boldsymbol{\psi}_1\\ \boldsymbol{\psi}_2\\ \ldots\\ \boldsymbol{\psi}_N\end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t^{\top}\hat{\boldsymbol{\Sigma}}_{\xi}^{-1}\boldsymbol{\psi}_t. \tag{A.41}$$

Because $\boldsymbol{\xi}_t$ is i.i.d. over time, the strong law of large numbers implies that the sample variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\xi}$ converges to the true $\boldsymbol{\Sigma}_{\xi}$ almost surely as T tends to infinity. Because $\boldsymbol{\xi}_t$ is the assets' fundamental returns and $\mathrm{var}(\mathbf{R}_0)$ is the fundamental-risk matrix, we have $\boldsymbol{\Sigma}_{\xi} = \mathrm{var}(\mathbf{R}_0)$. Therefore, in the limit of T tending to infinity, we have almost surely,

$$\hat{\mathbf{a}}^{\top}\frac{\mathrm{var}(\hat{\mathbf{a}})^{-1}}{T}\hat{\mathbf{a}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t^{\top}\mathrm{var}(\mathbf{R}_0)^{-1}\boldsymbol{\psi}_t. \tag{A.42}$$

Next, we transform the squared MPIR in (23). 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 in the main text. We now define $\check{\boldsymbol{\delta}}_n = (\check{\delta}_{n,1}, \check{\delta}_{n,2}, \ldots, \check{\delta}_{n,T})^{\top}$. Using equations (A.33) and (A.40), we have

$$\check{\boldsymbol{\delta}}_n = \mathbf{e}(\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\boldsymbol{\xi}_n + \sum_{k=1}^{K}\hat{\lambda}_k\left(\sum_{m=1}^{N}v_{n,m}\tilde{b}_{m,k}\right)\tilde{\mathbf{q}}_k = \boldsymbol{\psi}_n + \sum_{k=1}^{K}\hat{\lambda}_k\left(\sum_{m=1}^{N}v_{n,m}\tilde{b}_{m,k}\right)\tilde{\mathbf{q}}_k, \tag{A.43}$$

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

$$\check{\boldsymbol{\delta}}_t = \boldsymbol{\psi}_t + \mathrm{var}(\mathbf{R}_0)\sum_{k=1}^{K}\hat{\lambda}_k\tilde{q}_{k,t}\tilde{\mathbf{b}}_k. \tag{A.44}$$

The realized squared MPIR at time t is

$$\begin{aligned}
&\check{\boldsymbol{\delta}}_t^{\top}\mathrm{var}(\mathbf{R}_0)^{-1}\check{\boldsymbol{\delta}}_t \\
=&\left(\boldsymbol{\psi}_t^{\top}+\sum_{k=1}^{K}\hat{\lambda}_k\tilde{q}_{k,t}\tilde{\mathbf{b}}_k^{\top}\mathrm{var}(\mathbf{R}_0)\right)\mathrm{var}(\mathbf{R}_0)^{-1}\left(\boldsymbol{\psi}_t+\mathrm{var}(\mathbf{R}_0)\sum_{k=1}^{K}\hat{\lambda}_k\tilde{q}_{k,t}\tilde{\mathbf{b}}_k\right) \\
=&\boldsymbol{\psi}_t^{\top}\mathrm{var}(\mathbf{R}_0)^{-1}\boldsymbol{\psi}_t+2\boldsymbol{\psi}_t^{\top}\sum_{k=1}^{K}\hat{\lambda}_k\tilde{q}_{k,t}\tilde{\mathbf{b}}_k+\left(\sum_{k=1}^{K}\hat{\lambda}_k\tilde{q}_{k,t}\tilde{\mathbf{b}}_k^{\top}\right)\mathrm{var}(\mathbf{R}_0)\sum_{k=1}^{K}\hat{\lambda}_k\tilde{q}_{k,t}\tilde{\mathbf{b}}_k \\
=&\boldsymbol{\psi}_t^{\top}\mathrm{var}(\mathbf{R}_0)^{-1}\boldsymbol{\psi}_t+2\boldsymbol{\psi}_t^{\top}\sum_{k=1}^{K}\hat{\lambda}_k\tilde{q}_{k,t}\tilde{\mathbf{b}}_k+\sum_{k=1}^{K}\hat{\lambda}_k^2\tilde{q}_{k,t}^2, \qquad \text{(A.45)}
\end{aligned}$$

where, in the last step, we use the orthogonalization $\tilde{\mathbf{B}}^{\top}\mathrm{var}(\mathbf{R}_0)\tilde{\mathbf{B}}=\mathbf{I}_K$. The time-series average is

$$\begin{aligned}
\check{\theta}^2=&\frac{1}{T}\sum_{t=1}^{T}\check{\boldsymbol{\delta}}_n^{\top}\mathrm{var}(\mathbf{R}_0)^{-1}\check{\boldsymbol{\delta}}_n \\
=&\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t^{\top}\mathrm{var}(\mathbf{R}_0)^{-1}\boldsymbol{\psi}_t+\frac{2}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t^{\top}\sum_{k=1}^{K}\hat{\lambda}_k\tilde{q}_{k,t}\tilde{\mathbf{b}}_k+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\hat{\lambda}_k^2\tilde{q}_{k,t}^2. \qquad \text{(A.46)}
\end{aligned}$$

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

$$\begin{aligned}
\sum_{t=1}^{T}\tilde{q}_{k,t}\psi_{n,t}=&\sum_{t=1}^{T}\tilde{q}_{k,t}\mathbf{e}_t(\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\boldsymbol{\xi}_n \\
=&\left(\sum_{t=1}^{T}\tilde{q}_{k,t}e_{1,t},\sum_{t=1}^{T}\tilde{q}_{k,t}e_{2,t},\ldots,\sum_{t=1}^{T}\tilde{q}_{k,t}e_{N,t}\right)(\mathbf{e}^{\top}\mathbf{e})^{-1}\mathbf{e}^{\top}\boldsymbol{\xi}_n=0. \qquad \text{(A.47)}
\end{aligned}$$

Therefore, we know that

$$\check{\theta}^2=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t^{\top}\mathrm{var}(\mathbf{R}_0)^{-1}\boldsymbol{\psi}_t+\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\hat{\lambda}_k^2\tilde{q}_{k,t}^2. \qquad \text{(A.48)}$$

A similar calculation gives

$$\theta^2=\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\hat{\lambda}_k^2\tilde{q}_{k,t}^2. \qquad \text{(A.49)}$$

Therefore, we have

$$\check{\theta}^2\quad\theta^2=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t^{\top}\mathrm{var}(\mathbf{R}_0)^{-1}\boldsymbol{\psi}_t. \qquad \text{(A.50)}$$

Using (A.42), we have almost surely in the limit of T tending to infinity,

$$\hat{\mathbf{a}}^{\top}\frac{\text{var}(\hat{\mathbf{a}})^{-1}}{T}\hat{\mathbf{a}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\psi}_t^{\top}\text{var}(\mathbf{R}_0)^{-1}\boldsymbol{\psi}_t = \check{\theta}^2 \quad \theta^2. \tag{A.51}$$

B Technical Details Omitted in the Main Text

We present technical details omitted in the main text.

B.1 Illustration of Portfolio-Flow Construction

We illustrate portfolio-flow construction. The following examples show why the pseudoinverse of portfolio weights instead of the transpose constructs the correct portfolio flows.

Example 1. We consider two stocks with flows equal to \$1 and \$3. A market portfolio equally weights the two stocks. The portfolio weight matrix is therefore

$$\mathbf{W} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}. \tag{B.1}$$

The pseudoinverse recovers the correct magnitude of market portfolio flow

$$(\mathbf{W}^{\top}\mathbf{W})^{-1}\mathbf{W}^{\top}\begin{pmatrix} 1 \\ 3 \end{pmatrix} = 4, \tag{B.2}$$

because \$4 flows into the market portfolio allocate \$2 each to the two stocks. This is the best OLS approximation to the observed stock-level flows \$1 and \$3. In contrast, the transpose leads to an incorrect measure of portfolio flows,

$$\mathbf{W}^{\top}\begin{pmatrix} 1 \\ 3 \end{pmatrix} = 2. \tag{B.3}$$

The market flow is clearly \$4 not \$2. In this example, the market portfolio flow calculated using the transpose is off by a scaling.

Example 2. We consider three stocks with flows equal to \$3, \$0, and -\$3. There are two long-short portfolios. The first portfolio longs stock one and shorts stock two. The second portfolio longs stock one and shorts stocks two and three in equal weights. The portfolio

weight matrix is therefore

$$\mathbf{W} = \begin{pmatrix} 1 & 1 \\ 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}. \tag{B.4}$$

The pseudoinverse recovers the correct magnitude of flow into the two portfolios,

$$(\mathbf{W}^{\top}\mathbf{W})^{-1}\mathbf{W}^{\top} \begin{pmatrix} 3 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}. \tag{B.5}$$

Note that \$3 outflow from portfolio one takes \$3 out of stock one and puts \$3 into stock two, and \$6 flow into portfolio two puts \$6 into stock one and takes \$3 each out of stocks two and three. The sum is exactly the observed stock-level flows \$3, \$0, and -\$3. In contrast, the transpose leads to an incorrect measure of portfolio flows,

$$\mathbf{W}^{\top} \begin{pmatrix} 3 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 4.5 \end{pmatrix}. \tag{B.6}$$

In this example, the portfolio-one flow calculated using the transpose has the wrong sign.

B.2 Details for Orthogonalization

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 $\mathbf{q} = (q_1, q_2, \ldots, q_K)^{\top}$. Our goal is to find some $K \times K$ invertible matrix $\mathbf{O}$ such that

$$\mathbf{O}^{\top}\mathbf{B}^{\top}\mathrm{var}(\mathbf{R}_0)\mathbf{B}\mathbf{O} = \mathbf{I}_K, \tag{B.7}$$

$$\mathbf{O}\mathbf{\Pi}\mathbf{O}^{\top} = \mathrm{var}(\mathbf{q}) = \frac{1}{T}\begin{pmatrix} \sum_{t=1}^{T} q_{1,t}^2 & \sum_{t=1}^{T} q_{1,t}q_{2,t} & \sum_{t=1}^{T} q_{1,t}q_{K,t} \\ \sum_{t=1}^{T} q_{2,t}q_{1,t} & \sum_{t=1}^{T} q_{2,t}^2 & \sum_{t=1}^{T} q_{2,t}q_{K,t} \\ \vdots & \vdots & \vdots \\ \sum_{t=1}^{T} q_{K,t}q_{1,t} & \sum_{t=1}^{T} q_{K,t}q_{2,t} & \sum_{t=1}^{T} q_{K,t}^2 \end{pmatrix}, \tag{B.8}$$

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

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

$$\mathbf{B}^{\top}\mathrm{var}(\mathbf{R}_0)\mathbf{B} = \mathbf{U}^{\top}\mathbf{U}, \tag{B.9}$$

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

$$\left(\mathbf{U}\text{var}(\mathbf{q})\mathbf{U}^{>}\right)\mathbf{G} = \mathbf{G}\mathbf{\Pi}, \tag{B.10}$$

where $\mathbf{\Pi} = \text{diag}(\pi_1, \pi_2, \ldots, \pi_K)$, and $\mathbf{G}$ is an orthogonal $K \quad K$ matrix satisfying $\mathbf{G}^{\ 1} = \mathbf{G}^{>}$.

We claim that $\mathbf{O} = \mathbf{U}^{\ 1}\mathbf{G}$ satisfies the orthogonalization conditions (B.7) and (B.8). First, we see that

$$\mathbf{O}^{>}\mathbf{B}^{>}\text{var}(\mathbf{R}_0)\mathbf{B}\mathbf{O} = \mathbf{G}^{>}(\mathbf{U}^{>})^{\ 1}\mathbf{U}^{>}\mathbf{U}\mathbf{U}^{\ 1}\mathbf{G} = \mathbf{I}_K. \tag{B.11}$$

Second, we have by (B.10),

$$\mathbf{U}\text{var}(\mathbf{q})\mathbf{U}^{>}\mathbf{U}\mathbf{O} = \mathbf{U}\mathbf{O}\mathbf{\Pi}. \tag{B.12}$$

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

$$\text{var}(\mathbf{q})\mathbf{U}^{>}\mathbf{U}\mathbf{O} = \mathbf{O}\mathbf{\Pi}. \tag{B.13}$$

Note that

$$\mathbf{O}^{>}\mathbf{U}^{>}\mathbf{U}\mathbf{O} = \mathbf{G}^{>}\mathbf{G} = \mathbf{I}_K. \tag{B.14}$$

Therefore, we have

$$\text{var}(\mathbf{q})(\mathbf{O}^{>})^{\ 1} = \mathbf{O}\mathbf{\Pi}, \tag{B.15}$$

which proves (B.8).

We now use $\mathbf{O}$ to construct the orthogonalized factor portfolios $\tilde{b}_{n,k}$ and factor flows $\tilde{q}_{k,t}$,

$$\mathbf{B}\mathbf{O} = \tilde{\mathbf{B}} = (\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \ldots, \tilde{\mathbf{b}}_K) = f\tilde{b}_{n,k}g, \tag{B.16}$$

$$\mathbf{O}^{\ 1}\begin{pmatrix} q_{1,t} \\ q_{2,t} \\ \\ q_{K,t} \end{pmatrix} = \begin{pmatrix} \tilde{q}_{1,t} \\ \tilde{q}_{2,t} \\ \\ \tilde{q}_{K,t} \end{pmatrix}, \tag{B.17}$$

for all times $t = 1, 2, \ldots, T$. The orthogonalized factor portfolios $\tilde{\mathbf{B}}$ form an $N \quad K$ matrix, with the k-th factor being $\tilde{\mathbf{b}}_k$ and the (n,k)-th entry being $\tilde{b}_{n,k}$. By (B.7) and (B.8), the K orthogonalized factors have uncorrelated fundamental returns and uncorrelated flows,

$$\text{var}(\tilde{\mathbf{B}}^{>}\mathbf{R}_0) \text{ is an identity matrix } \mathbf{I}_K, \tag{B.18}$$

$$\text{var}(\tilde{\mathbf{q}}) \text{ is a diagonal matrix } \mathbf{\Pi} = \text{diag}(\pi_1, \pi_2, \ldots, \pi_K). \tag{B.19}$$

The K orthogonalized factors have no cross impacts, so their $K \quad K$ price-of-risk matrix from (12) is a diagonal matrix $\tilde{\mathbf{\Lambda}} = \text{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_K)$. For the K orignal factors, the $K \quad K$ price-of-risk matrix in (14) is given by $\mathbf{\Lambda} = \mathbf{O}\tilde{\mathbf{\Lambda}}\mathbf{O}^{\ 1}$.

B.3 Arguments for the Exclusion Restriction of IV

We present detailed arguments for the exclusion restriction of the lagged-flow-difference IV.

We assume that the error term in the second-stage regression (26) follows

$$\xi_{n,t} = \psi_{n,t} \quad \frac{1}{H}\sum_{h=1}^{H}\sum_{k=1}^{3}\tilde{\lambda}_k \tilde{q}_{k,t\ \ h}\left(\sum_{m=1}^{N} v_{n,m}\tilde{b}_{m,k}\right), \tag{B.20}$$

where the first term $\psi_{n,t}$ is the fundamental return of asset n in month t and the second term models the flow-induced price reversion. The second term assumes that the price impacts of flows $\tilde{q}_{k,t\ \ 1}, \tilde{q}_{k,t\ \ 2}, \ldots, \tilde{q}_{k,t\ \ H}$ of the past H months revert to zero at a linear rate.

We claim that $\tilde{q}_{k,t\ \ 1} \quad \tilde{q}_{k,t\ \ H}$ for $k = 1, 2, 3$ is a reasonable IV that satisfies the exclusion restriction $\mathsf{E}[(\tilde{q}_{k,t\ \ 1} \quad \tilde{q}_{k,t\ \ H})\xi_{n,t}] = 0$. To see this, note that for $\tilde{q}_{1,t\ \ 1} \quad \tilde{q}_{1,t\ \ H}$, we have

$$\begin{aligned}
&\mathsf{E}[(\tilde{q}_{1,t\ \ 1} \quad \tilde{q}_{1,t\ \ H})\xi_{n,t}] \\
=&\,\mathsf{E}[(\tilde{q}_{1,t\ \ 1} \quad \tilde{q}_{1,t\ \ H})\psi_{n,t}] \quad \sum_{h=1}^{H}\sum_{k=2}^{3}\tilde{\lambda}_k \mathsf{E}[(\tilde{q}_{1,t\ \ 1} \quad \tilde{q}_{1,t\ \ H})\tilde{q}_{k,t\ \ h}]\left(\sum_{m=1}^{N} v_{n,m}\tilde{b}_{m,k}\right) \\
&\sum_{h=1}^{H}\tilde{\lambda}_1 \mathsf{E}[(\tilde{q}_{1,t\ \ 1} \quad \tilde{q}_{1,t\ \ H})\tilde{q}_{1,t\ \ h}]\left(\sum_{m=1}^{N} v_{n,m}\tilde{b}_{m,k}\right).
\end{aligned} \tag{B.21}$$

We expect the first term $\mathsf{E}[(\tilde{q}_{1,t\ \ 1} \quad \tilde{q}_{1,t\ \ H})\psi_{n,t}] = 0$ as long as past flows do not predict future fundamental returns. This is a reasonable assumption because our flows represent the non-discretionary component of mutual fund flows that are induced by retail investors. We expect the second term $\mathsf{E}[(\tilde{q}_{1,t\ \ 1} \quad \tilde{q}_{1,t\ \ H})\tilde{q}_{k,t\ \ h}] = 0$ for $k = 2, 3$ and $h = 1, 2, \ldots, H$, because factors 1 and k are orthogonalized such that their flows are uncorrelated. For the third term, note that

$$\sum_{h=1}^{H}\mathsf{E}[(\tilde{q}_{1,t\ \ 1} \quad \tilde{q}_{1,t\ \ H})\tilde{q}_{1,t\ \ h}] = \sum_{h=1}^{H}\mathsf{E}[\tilde{q}_{1,t\ \ 1}\tilde{q}_{1,t\ \ h} \quad \tilde{q}_{1,t\ \ H+h\ \ 1}\tilde{q}_{1,t\ \ H}], \tag{B.22}$$

where both $\mathsf{E}[\tilde{q}_{1,t\ \ 1}\tilde{q}_{1,t\ \ h}]$ and $\mathsf{E}[\tilde{q}_{1,t\ \ H+h\ \ 1}\tilde{q}_{1,t\ \ H}]$ are the $h \quad 1$ order autocovariance of the factor-1 flow, so are equal to each other. To sum up, it is reasonable to expect that $\mathsf{E}[(\tilde{q}_{k,t\ \ 1} \quad \tilde{q}_{k,t\ \ H})\xi_{n,t}] = 0$ for $k = 1, 2, 3$, which means that $\tilde{q}_{k,t\ \ 1} \quad \tilde{q}_{k,t\ \ H}$ is a valid IV.

B.4 Details for MPIR strategy

We provide details on constructing the MPIR strategy and combining it with fundamental investing.

First, using equation (29), we define the staggered MPIR portfolio in month s as

$$\tilde{\mathbf{c}}_s^{ftg} = \sum_{k=1}^{3} \tilde{\lambda}_k^{ftg} \bar{q}_{k,s}^{ftg} \tilde{\mathbf{b}}_k^{ftg}, \tag{B.23}$$

where the superscript ftg is because $\tilde{\mathbf{c}}_s^{ftg}$ is estimated using the training window up to month t. The term $\tilde{\lambda}_k^{ftg}$ is the estimated price of risk of orthogonalized factor k, $\bar{q}_{k,s}^{ftg}$ is the average flow into factor k over past six months $s \quad 5, s \quad 4, \ldots, s$, and $\tilde{\mathbf{b}}_k^{ftg}$ is the mimicking portfolio weights of factor k.

Second, we define the model-implied price impact ratio and normalized price reversion of the staggered MPIR portfolio $\tilde{\mathbf{c}}_s^{ftg}$. By equation (30), the price impact ratio in month s is

$$\kappa_s^{ftg} := \frac{(\tilde{\mathbf{c}}_s^{ftg})^\top \mathrm{var}(\mathbf{R}_0)^{ftg} \sum_{k=1}^{3} \tilde{\lambda}_k^{ftg} \bar{q}_{k,s}^{ftg} \tilde{\mathbf{b}}_k^{ftg}}{\sqrt{(\tilde{\mathbf{c}}_s^{ftg})^\top \mathrm{var}(\mathbf{R}_0)^{ftg} \tilde{\mathbf{c}}_s^{ftg}}} = \sqrt{\sum_{k=1}^{3} \left(\tilde{\lambda}_k^{ftg} \bar{q}_{k,s}^{ftg}\right)^2}, \tag{B.24}$$

where we use the orthogonalization $(\tilde{\mathbf{b}}_k^{ftg})^\top \mathrm{var}(\mathbf{R}_0)^{ftg} \tilde{\mathbf{b}}_k^{ftg} = 1$ and $(\tilde{\mathbf{b}}_k^{ftg})^\top \mathrm{var}(\mathbf{R}_0)^{ftg} \tilde{\mathbf{b}}_l^{ftg} = 0$ for $k \neq l$. Similarly, the normalized price reversion of $\tilde{\mathbf{c}}_s^{ftg}$ in month $s+1$ is

$$\breve{\kappa}_s^{ftg} := \frac{(\tilde{\mathbf{c}}_s^{ftg})^\top \mathbf{r}_{s+1}}{\sqrt{(\tilde{\mathbf{c}}_s^{ftg})^\top \mathrm{var}(\mathbf{R}_0)^{ftg} \tilde{\mathbf{c}}_s^{ftg}}} = \frac{(\tilde{\mathbf{c}}_s^{ftg})^\top \mathbf{r}_{s+1}}{\sqrt{\sum_{k=1}^{3} \left(\tilde{\lambda}_k^{ftg} \bar{q}_{k,s}^{ftg}\right)^2}}, \tag{B.25}$$

where $\mathbf{r}_{s+1}$ is the demeaned[29] 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

$$\mathrm{REV}^{ftg} = \frac{\sum_{s=6}^{t \quad 1} \breve{\kappa}_s^{ftg}}{\sum_{s=6}^{t \quad 1} \kappa_s^{ftg}}. \tag{B.26}$$

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

Fourth, the staggered MPIR portfolio that we trade against in month $t+1$ is

$$\tilde{\mathbf{c}}_t = \min\left(\max\left(\mathrm{REV}^{ftg}, 0\right), 1\right) \tilde{\mathbf{c}}_t^{ftg}. \tag{B.27}$$

[29] We remove the time-series mean of $\mathbf{r}_{s+1}$ for $s = 6, \ldots, t \quad 1$ by sample averages.

The first term $\min\left(\max\left(\text{REV}^{ftg}, 0\right), 1\right)$ is the shrinkage to parsimoniously account for the potential over-estimation of the price of risk $\tilde{\lambda}_k^{ftg}$ in the OLS regression. If we do not apply the shrinkage, we then simply trade against the staggered portfolio $\tilde{\mathbf{c}}_t^{ftg}$.

Fifth, the MPIR strategy's excess return in month $t+1$ is defined as

$$\tilde{r}_{t+1} = \quad (\tilde{\mathbf{c}}_t)^{\top}(\mathbf{r}_{t+1} \quad r_{F,t+1}), \tag{B.28}$$

where $\mathbf{r}_{t+1}$ is the return of the 25 test assets in month $t+1$ and $r_{F,t+1}$ is the net risk-free rate in month $t+1$.

Lastly, we combine the MPIR strategy with fundamental investing using equation (32), with the theory-implied mixing ratio

$$w_{j,t} := 12 \max\left(\frac{\text{VOL}_j^{ftg}}{\text{SR}_j^{ftg}}, 0\right)(1 + r_{F,t+1}). \tag{B.29}$$

The multiplication by 12 normalizes the MPIR strategy return $\tilde{r}_{t+1}$ such that its monthly return volatility equals its price impact ratio. This is because of equation (B.24) and our measurement of $\text{var}(\mathbf{R}_0)$ by annualized returns. Moreover, SR_j^{ftg} is the Sharpe ratio of anomaly portfolio j and VOL_j^{ftg} is the return standard deviation. They are calculated using the same training windows as the MPIR estimation. We normalize the MRIP strategy by $\text{VOL}_j^{ftg}/\text{SR}_j^{ftg}$, rather than normalizing portfolio j by $\text{SR}_j^{ftg}/\text{VOL}_j^{ftg}$. We do so to avoid normalizing portfolio j by the inverse of volatility in the time series, where Moreira and Muir (2017) show that such normalization per se increases Sharpe ratio. Because the estimated Sharpe ratio SR_j^{ftg} could be negative, we bound the scaling by zero. That is, we leave $r_{j,t+1}$ unchanged if portfolio j has negative historical Sharpe ratio up to month t. The last term $1 + r_{F,t+1}$ is the gross riskfree rate in month $t+1$, which is implied by Theorem 1 and fully known at the start of month $t+1$.

C Construction and Cleaning of Mutual Fund Flows

We present the 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 return and total net assets (TNA) data at the share-class level from all funds. A mutual fund may include multiple share classes. We first drop observations with no valid CRSP identifier, `crsp_fundno`. A small number of fund share classes report multiple TNA in a given month. These are likely data duplication, 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 separately 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 extreme positive returns. We study the case in which a certain fund share has returns greater than 100% and has existed for more than one year.[30] We manually check the entire time series of each share class in this subsample. The majority of these extreme returns reflect misplaced decimal points, which confound returns in decimal and percentage formats. For these cases, we divide the erroneous returns by 100.

Second, we address extreme negative returns. Similarly, we obtain a subset of monthly fund observations, in which a certain fund share has existed for more than one year and has returns that are lower than 50%. With extreme 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. For this subsample, we manually check whether these extreme returns reflect data-input errors. For the cases with misplaced decimal points, we divide the erroneous returns by 100.

C.2 TNA Cleaning

Unlike many prior studies that construct percentage mutual fund flows, we study dollar-amount flows in order to preserve the relative magnitudes in the cross section. The mutual fund size distribution features a very long right tail. Winsorizing the extreme dollar-amount TNA likely removes valid large values together with 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 (24) at the share-class level.

2. We study the top and bottom 0.5% of all dollar flows.[31] We manually check the TNA time series of all share classes in this subsample. We identify several common error types:

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

[31] 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.

- 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.

We correct the misplaced decimal issue immediately. For funds suffering from the latter two problems, we obtain their TNA from Morningstar.[32] Morningstar's TNA data (`Net_Assets_ShareClass_Monthly`) suffer to a lesser extent from these issues than CRSP's TNA data. We arrive at this conclusion by further cross-checking other third-party vendors (e.g., Yahoo Finance and Bloomberg Terminal). Hence, when 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 make sure 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.[33] 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 (24). We obtain a fund's flows by adding up the flows of all share classes in the fund. The final sample contains 1,613,579 fund month observations.

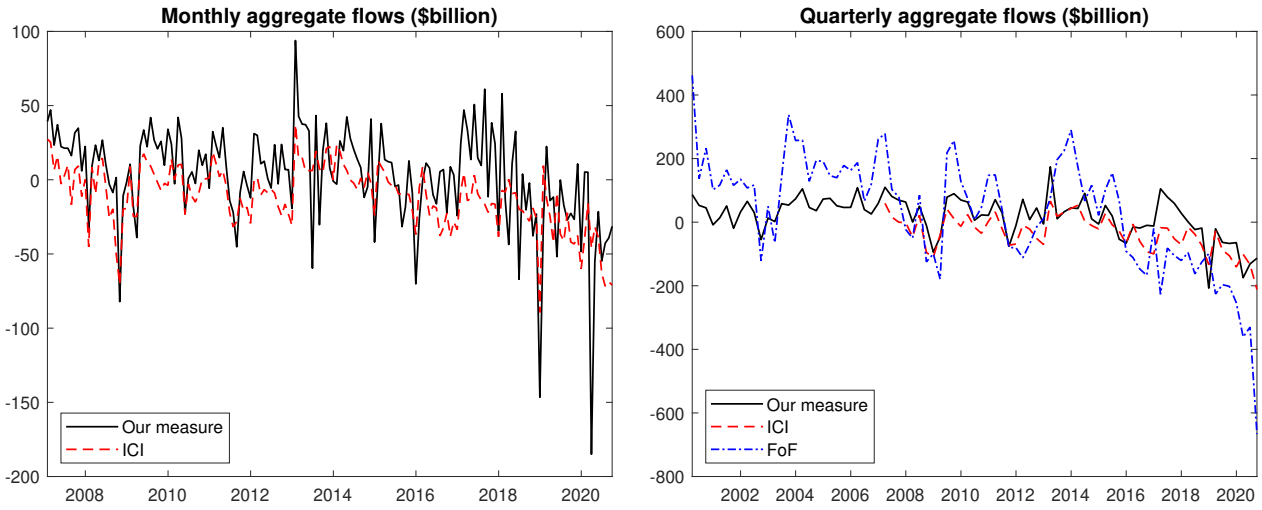
C.3 Cross-Validating the Data-Cleaning Procedure

We cross-validate our data-cleaning procedure. For each month, we compute the aggregate mutual fund flows in dollar amounts. We compare our aggregate flow measures with those from 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.

[32] We merge the CRSP and Morningstar databases using a fund's ticker.
[33] We merge the two databases via the linking table MFLINKS, which is provided by WRDS.

Figure C.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.

The FoF data (now known as "Financial Accounts of the United States - Z.1") are published quarterly by the 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 C.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. 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.

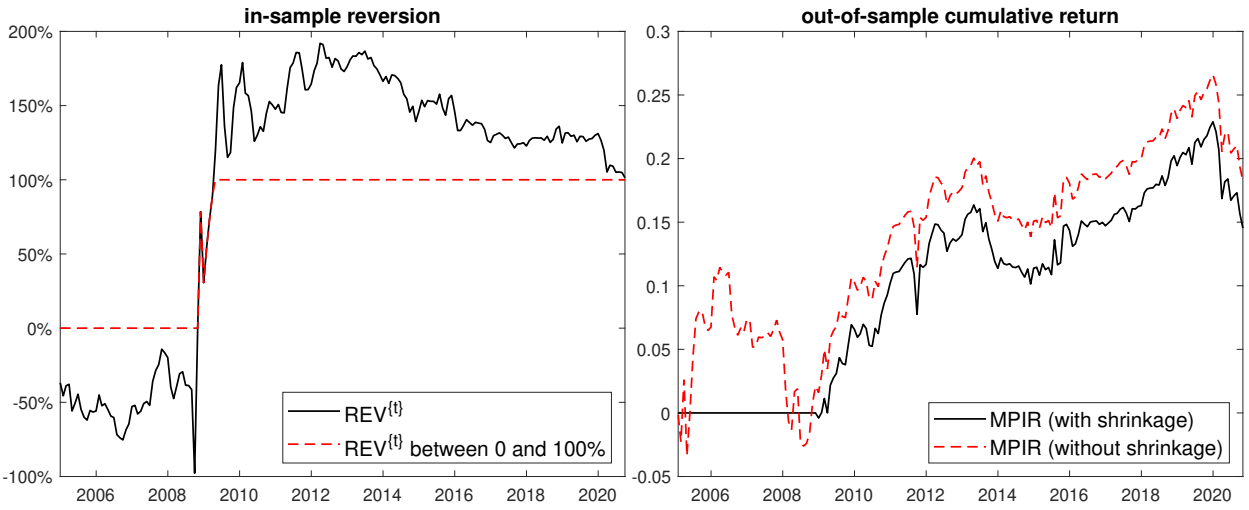
The differences in Figure C.1 between the three measures of aggregate flows likely reflect differences in mutual fund coverage. Specifically, the ICI flow tracks virtually all U.S. equity mutual funds that invest in both domestic and world equity markets.[34] The FoF flow, sourced from unpublished ICI 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 on mutual fund assets in common stock, preferred stock, and rights and warrants.[35] In comparison, our mutual fund sample contains U.S. mutual funds that are covered by CRSP. CRSP collects historical data from various sources.[36] Due to the nature of the data

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

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

[36]The sources include the Fund Scope Monthly Investment Company Magazine, the Investment Dealers

Figure D.1. MPIR strategy's in-sample reversion and out-of-sample return: day return



Note: We use day returns as the left-hand side in the second-stage regression to estimate the prices of risk and then construct the MPIR strategy. In the left panel, we show REV^{ftg}, the ratio of the average one-month-forward reversion to the average model-implied price impact of the staggered MPIR portfolios for each training window that starts in January 2000 and ends in month t. In the right panel, we show the MPIR strategy's cumulative log returns with and without applying the shrinkage for the out-of-sample period from January 2005 to September 2020.

collection process, CRSP's coverage is less complete than ICI's coverage.

D Additional Empirical Results

We provide additional empirical results.

D.1 Robustness Checks using Day Returns

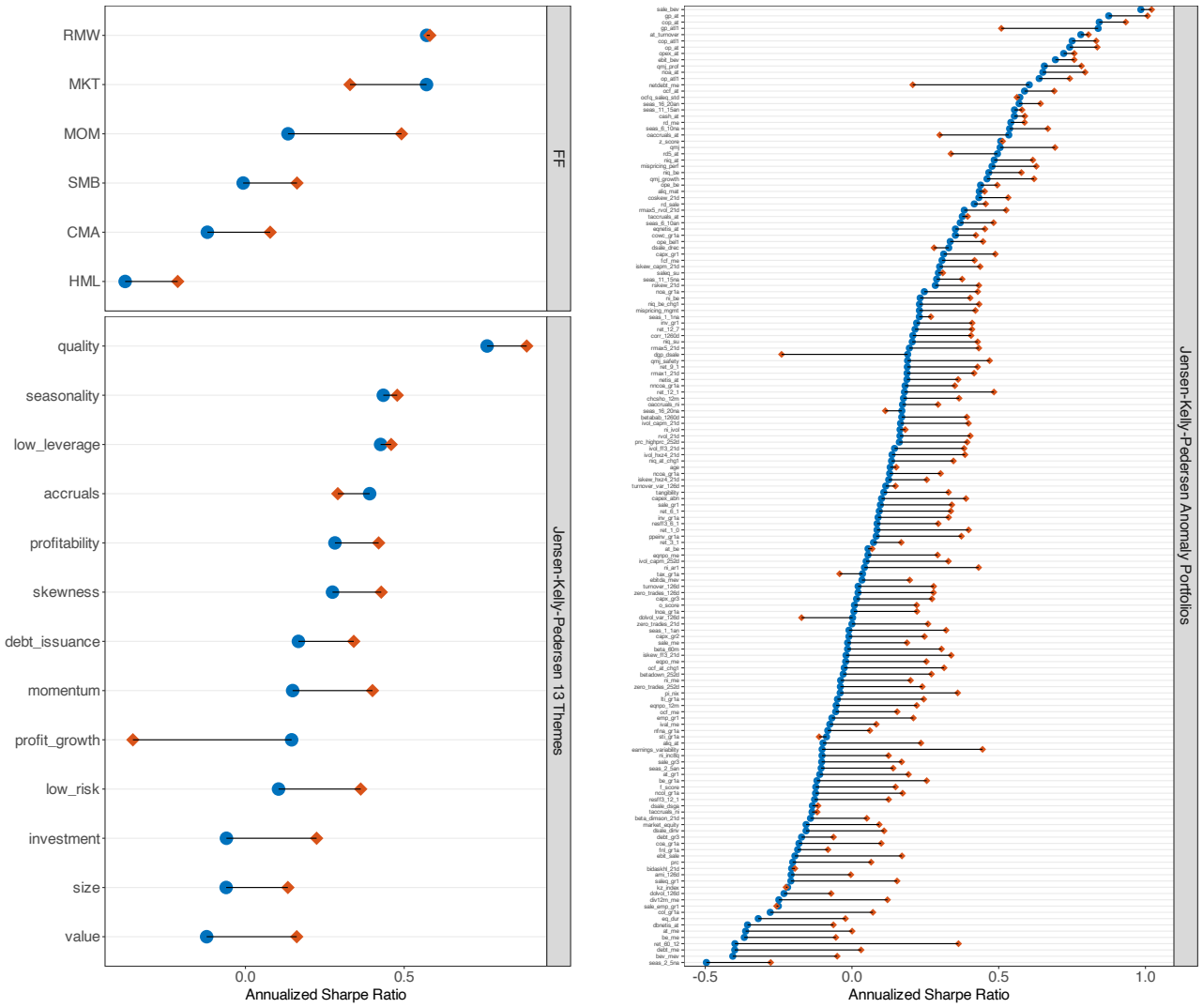
We provide robustness checks using the price of risk estimated from day returns.

Specifically, we use (open-to-close) day returns as the left-hand side in the second-stage regression to estimate the prices of risk and then construct MPIR strategy. We note that the only difference with the baseline specification is the estimation of prices of risk. When constructing the trading-against-MPIR strategy for month $t+1$, we still hold the position for the entire month, not just hold the position during day time.

Figure D.1 replicates Figure 4 in the main text and shows similar results. Specifically, the left panel shows that the reversion works better since the 2008 financial crisis. The right panel shows the MPIR strategy's cumulative out-of-sample log returns. The strategy's

Digest Mutual Fund Guide, Investor's Mutual Fund Guide, the United and Babson Mutual Fund Selector, and the Wiesenberger Investment Companies Annual Volumes.

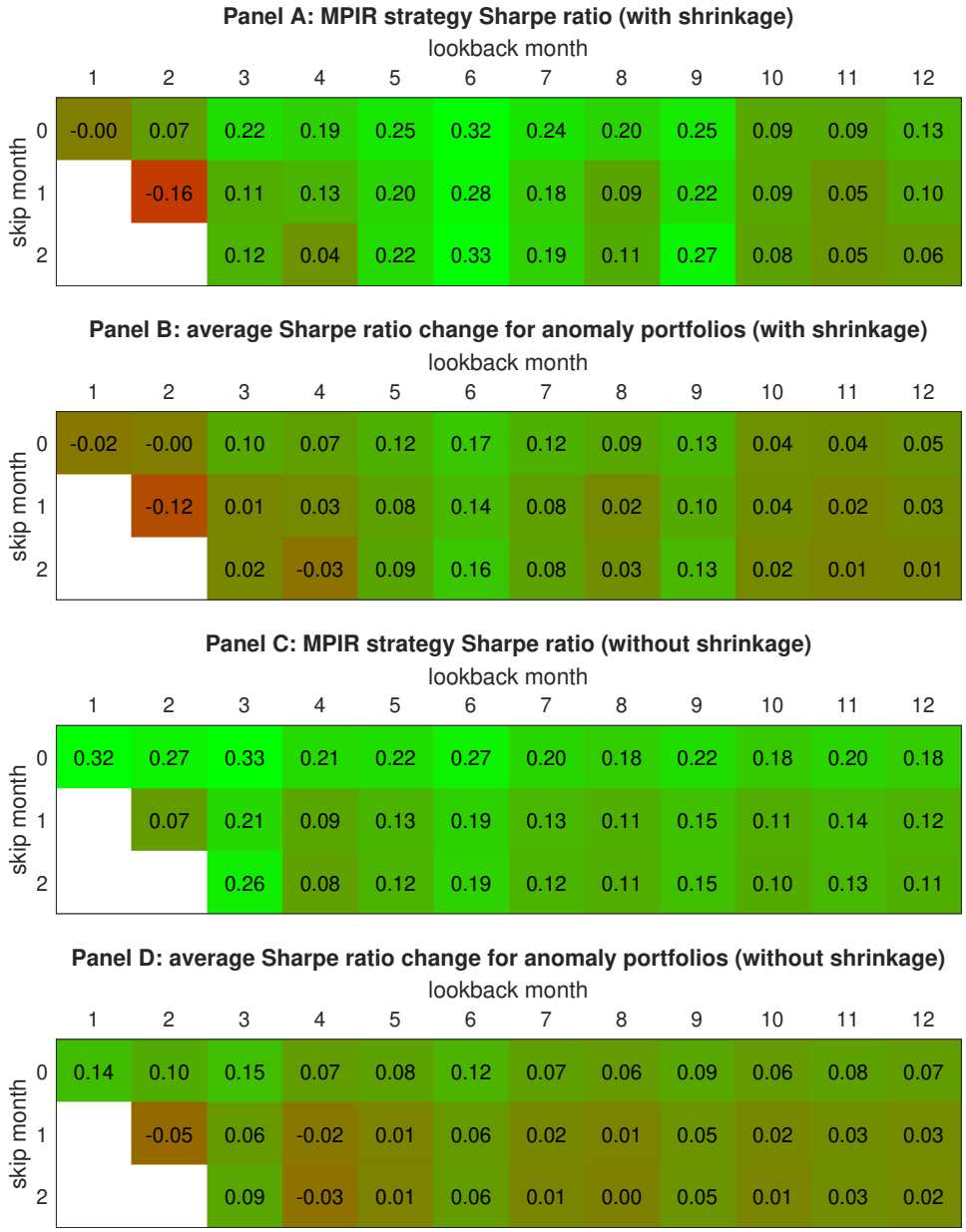
Figure D.2. MPIR increases out-of-sample Sharpe ratio of anomaly portfolios: day return



	# of factors	mean	std.	P5	P10	P25	median	P75	P90	P95
Sharpe ratio change	154	0.17	0.16	-0.06	0.01	0.09	0.19	0.26	0.33	0.36

Note: The blue dots represent the out-of-sample annualized Sharpe ratio of Jensen, Kelly, and Pedersen (2021) 154 portfolios, including Fama-French-Carhart six factors and a large list of other firm characteristics-based anomaly portfolios. The anomaly portfolios are also grouped by 13 themes. The red diamonds represent the Sharpe ratio of the portfolio that combines the original portfolio with our MPIR strategy. We use day returns as the left-hand side in the second-stage regression to estimate the prices of risk and then construct the MPIR strategy. The MPIR strategy trades against the MPIR portfolios of the past six months in equal weights and applies the shrinkage for the estimated prices of risk. In each month, we scale the MPIR strategy such that the combined portfolio allocates risk between the original anomaly and our MPIR strategy proportional to their respective Sharpe ratio and price impact ratio. The expanding estimation windows start with January 2000 to December 2004, and the out-of-sample testing period is January 2005 to September 2020.

Table D.1. Robustness to skip, lookback period, and shrinkage: day return



Note: We use day returns as the left-hand side in the second-stage regression to estimate the prices of risk and then construct the MPIR strategy. In the top two panels, we report the MPIR strategy Sharpe ratio and the average Sharpe ratio change for the anomaly portfolios for different skip and lookback months. Our baseline specification that trades against the past six months' flows corresponds to the $(0, 6)$ position. The top two panels apply the shrinkage and the bottom two panels do not. The out-of-sample testing period is January 2005 to September 2020.

annualized Sharpe ratios are 0.32 and 0.27, with and without applying the shrinkage. We also see that the MPIR strategy works better during the 2008 crisis.

Figure D.2 replicates Figure 5 in the main text and shows the similar results that MPIR strategy improves the investment outcomes of a broad spectrum of firm characteristics-based

anomaly portfolios. Of the 154 anomaly portfolios, 145 or 94% receive a positive increase in Sharpe ratio. The average change in Sharpe ratio is 0.17 and the median change is 0.19.

Table D.1 replicates Table 4 in the main text and shows the similar results that our MPIR strategy is robust to skip, lookback period, and shrinkage. Compared to Table 4, the overall investment performance of the MPIR strategy in Table D.1 is similar but slightly worsens.

D.2 Decomposition of the Maximum Price Impact Ratio

We use MPIR to decompose how the MKT, SMB, and HML factors explain the cross section of price impacts. Under our factor model, the cross section of price impacts is $\boldsymbol{\delta}_t = (\delta_{1,t}, \delta_{2,t}, \ldots, \delta_{N,t})^\top$, where $\delta_{n,t} = \sum_{k,l \in f\text{MKT,SMB,HML}g} \lambda_{k,l} q_{l,t} \left(\sum_{m=1}^{N} v_{n,m} b_{m,k} \right)$. This expression is obtained from (28) through rotation to the MKT, SMB, and HML factors.

For any subset of factors $L \subseteq f\text{MKT,SMB,HML}g$, we compute the corresponding price impact $\sum_{k \in f\text{MKT,SMB,HML}g} \sum_{l \in L} \lambda_{k,l} q_{l,t} \left(\sum_{m=1}^{N} v_{n,m} b_{m,k} \right)$, which shuts down the flow of factors other than L. Nevertheless, the L flows could still impact the prices of all of the MKT, SMB, and HML factors. We then compute the MPIR for each month t.

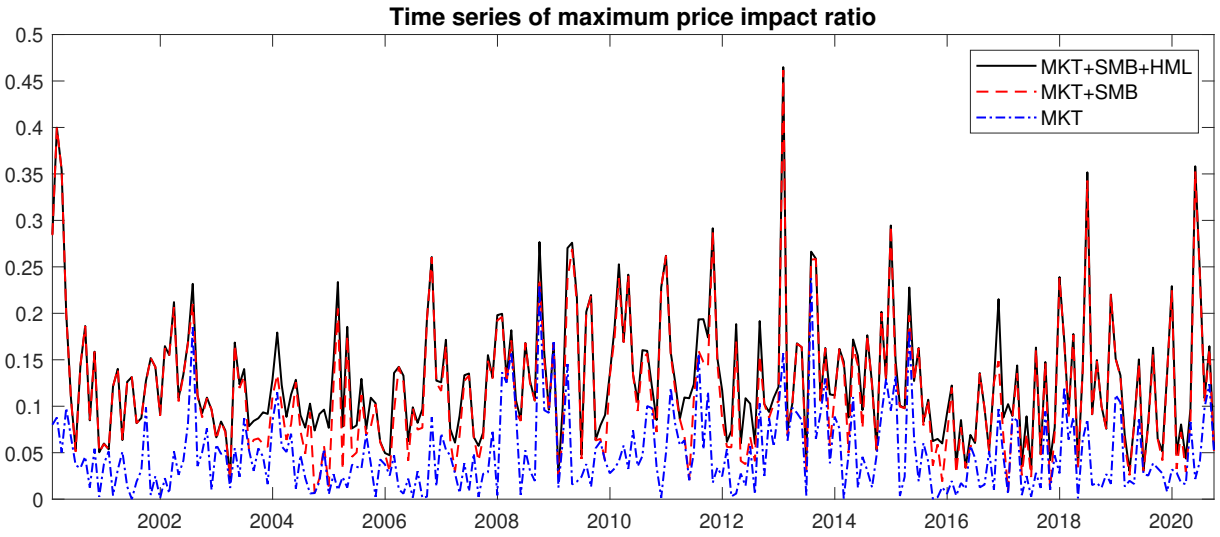
Figure D.3 shows the time series of MPIR for MKT+SMB+HML, MKT+SMB, and MKT. The full model (MKT+SMB+HML) exhibits a pretty volatile MPIR. In months with extreme flows, the MPIR can be as high as 0.4. That is, the MPIR portfolio's price dislocation is about 40% of its annual risk. Not all explanatory power for the cross section of price impacts arises from the MKT factor. The SMB factor also contributes greatly.

The bottom panel in Figure D.3 shows the time-series means of MPIR for all possible model combinations. The SMB factor provides the highest explanatory power for the cross section of price impacts, and the MKT and HML factors have roughly the same explanatory power. This finding may be surprising because MKT flow is about six times as volatile as the SMB and HML flows (Table 2). The reason is that the SMB and HML factors have higher prices of risk than the MKT factor (Table 3), and the quantity of risk induced by every dollar of flow also differs. Not all flows are created equal—the price of risk and the quantity of risk matter for the cross section of price impacts.

D.3 Raw and Anomaly Price Impacts

We quantify the raw price impacts created by the 25 assets' total flows and anomaly price impacts created by the 25 assets' idiosyncratic flows. We run the following asset-by-asset

Figure D.3. Decomposition of the maximum price impact ratio



time-series mean of the maximum price impact ratio

	MKT	SMB	HML	MKT+SMB	MKT+HML	SMB+HML	MKT+SMB+HML
mean	0.049	0.098	0.035	0.118	0.066	0.112	0.129
t-stats	(18.18)	(20.41)	(18.68)	(25.66)	(23.79)	(25.38)	(29.38)

Note: We compute the MPIR of a given subset of factors L fMKT,SMB,HMLg. We shut down the flow of factors other than L and compute the corresponding cross section of price impacts. We then compute the corresponding MPIR for each month t. The top figure plots the time series of the MPIR and the bottom table provides the time-series mean of the MPIR and the t-statistics of the mean.

time-series regressions,

$$r_{n,t} = \eta_n f_{n,t}/w_n + \epsilon_{n,t}, \tag{D.1}$$

$$r_{n,t} = \phi_n \left(\sum_{m \text{ neighboring to } n} f_{m,t}/w_m \right) \Big/ (\text{number of } m \text{ neighboring to } n) + \epsilon_{n,t}. \tag{D.2}$$

The first regression studies whether an asset's total flow $f_{n,t}$ impacts its own price. The second regression studies whether the average flow into neighboring assets impacts an asset's price. We divide by the extra market capitalization weight w_n in the regressions, such that our estimated multipliers η_n and ϕ_n are comparable to the standard price impact multipliers.

Similarly, to identify the anomaly price impacts created by idiosyncratic flows, we run the same regressions (D.1) and (D.2), but substitute total flows $f_{n,t}$ by idiosyncratic flows $e_{n,t}$. Note that the regression (31) in the main text also contains the factor component

Table D.2. Raw price impacts and anomaly price impacts

	raw price impact					raw cross impact				
	Low	2	3	4	High	Low	2	3	4	High
	regression R^2					regression R^2				
Small	6.62%	7.30%	9.34%	9.97%	9.05%	8.20%	8.74%	10.76%	10.33%	11.39%
2	8.78%	8.99%	10.34%	9.97%	8.21%	7.82%	9.18%	11.54%	10.12%	8.82%
3	7.72%	9.92%	9.99%	9.05%	7.97%	8.96%	11.05%	10.05%	11.28%	9.54%
4	12.53%	10.24%	8.40%	9.78%	6.92%	8.48%	9.77%	10.08%	7.89%	8.94%
Big	3.43%	3.01%	2.54%	1.07%	6.01%	6.13%	8.72%	3.92%	5.22%	5.17%
	multiplier η					multiplier ϕ				
Small	16.33	12.96	10.94	10.87	11.85	16.95	13.86	12.68	11.37	13.40
2	14.03	10.61	11.00	10.76	11.76	15.16	12.77	12.70	12.20	14.20
3	13.74	12.12	12.39	11.71	12.86	16.32	14.49	13.17	14.66	14.28
4	21.75	16.87	16.12	17.49	12.91	20.22	18.67	18.92	16.05	17.99
Big	16.66	12.62	10.63	7.14	14.77	16.92	19.86	13.38	14.65	14.58
	$t(\eta)$					$t(\phi)$				
Small	4.37	4.44	4.18	4.86	4.17	4.88	4.84	4.84	5.08	4.70
2	4.73	4.46	4.43	5.01	4.18	4.67	4.50	5.24	4.84	4.59
3	4.14	4.93	5.15	4.92	4.64	4.42	5.29	4.79	5.33	4.94
4	5.15	4.53	3.90	4.45	5.01	4.28	4.27	4.51	4.05	4.97
Big	2.30	2.28	2.24	1.29	4.06	3.31	4.15	2.45	3.73	3.13
	anomaly price impact					**anomaly cross impact**				
	regression R^2					regression R^2				
Small	1.04%	0.98%	1.80%	2.17%	1.48%	1.58%	1.36%	2.46%	2.18%	2.38%
2	2.30%	2.24%	1.60%	1.52%	0.76%	1.76%	2.38%	2.32%	1.62%	1.28%
3	2.43%	2.19%	2.29%	1.07%	0.99%	3.79%	2.71%	2.21%	2.00%	1.45%
4	7.36%	5.18%	3.66%	4.03%	0.63%	3.19%	3.91%	4.54%	1.87%	1.13%
Big	0.19%	0.28%	0.09%	0.86%	1.16%	3.23%	2.76%	0.36%	1.77%	0.22%
	multiplier $\check{\eta}$					multiplier $\check{\phi}$				
Small	8.96	6.90	6.98	7.58	6.93	11.88	8.71	9.56	8.23	10.00
2	11.26	8.73	6.75	6.83	5.93	12.22	10.90	9.67	8.26	9.26
3	12.27	8.81	9.69	6.20	7.43	18.00	12.44	10.17	11.09	10.21
4	25.65	18.54	15.80	19.52	6.78	21.84	20.63	24.62	15.36	13.02
Big	10.47	-10.14	5.02	-18.14	12.59	26.63	30.59	10.63	22.79	7.13
	$t(\check{\eta})$					$t(\check{\phi})$				
Small	1.52	1.54	1.95	2.56	1.95	1.76	1.68	2.58	2.61	2.53
2	2.19	2.17	2.02	2.23	1.45	1.85	2.29	2.78	2.27	2.20
3	2.24	2.51	2.79	1.79	1.68	2.59	2.78	2.63	2.69	1.99
4	4.19	3.12	2.59	2.79	1.34	2.62	2.84	3.09	2.14	1.64
Big	0.56	-0.69	0.46	-1.42	1.19	2.73	2.42	0.90	1.88	0.75

Note: The top half presents the R^2, point estimates of multipliers, and t-statistics for raw price impact regression (D.1) and cross impact regression (D.2). The 5 5 assets are sorted based on size (small to big) and book-to-market equity (low to high). To obtain the raw price impact, we run time-series regressions of each asset's demeaned return on its own flow. To obtain the raw cross impact, we run time-series regressions of each asset's demeaned return on its neighboring assets' average flow. The t-statistics are calculated from heteroskedasticity-robust standard errors. The bottom half presents the results for anomaly impact regressions, where we use idiosyncratic flows, the residuals from the first-stage flow regression (25), as right-hand sides. We bootstrap 10,000 times to obtain the t-statistics for anomaly impact regressions.

$\sum_{k=1}^{3} \tilde{\lambda}_k \tilde{q}_{k,t} \left(\sum_{m=1}^{N} v_{n,m} \tilde{b}_{m,k} \right)$. We prove in Appendix A.4 that ignoring the factor component does not change the estimated anomaly price impacts, which justifies the substitution of total flows $f_{n,t}$ in regressions (D.1) and (D.2) by idiosyncratic flows $e_{n,t}$.

Table D.2 presents the raw impact and anomaly impact regression results. For an average 5 5 asset, raw flows explain about 10% of variations of its own price and of neighboring assets' price. Idiosyncratic flows explain about 3% of return variations. Consistently, our factor model regression (26) has an R^2 of about 7%. Our three-price-of-risk model explains about 70% of cross-sectional price impacts.

The point estimates of raw impact and anomaly impact multipliers are between 10 and 20. The anomaly impact multipliers are generally smaller in magnitudes than the raw impact multipliers, though the differences lack statistical significance. Our multiplier estimates are on the larger side of literature estimates.[37] One reason is that flows at the 5 5 asset level may be informed or chase concurrent returns, leading to potentially biased multiplier estimates. The non-reverting price impacts of the factor-plus-idiosyncratic-flow MPIR portfolio presented in Figure 3 are consistent with this interpretation. That being said, our factor-model price impacts and MPIR strategy depend only on factor flows and do not use the raw or anomaly impact multipliers of the 5 5 assets. The IV estimation and the shrinkage MPIR strategy presented in the main text largely avoid these endogeneity concerns.

D.4 Other Empirical Results

In Table D.3, we report the IV first-stage regression results, where each orthogonalized factor's flow is regressed on its concurrent night return and the difference between one-month and six-month lagged flows. The flow into orthogonalized factor 1, which corresponds to the orthogonalized MKT direction, strongly chases concurrent night returns. The flow into orthogonalized factors 2 and 3, which corresponds to the orthogonalized SMB and HML direction, shows stronger effects to lagged flow differences. The F-statistics for the IV first-stage regression are around six for all three factors.

In Table D.4 we direct run (27) as a panel regression without imposing the IUF restriction, such that the 3 3 prices of risk $\lambda_{k,l}$ are 9 free parameters. Compared to the the IUF version presented in panel B of Table 3, the prices of risk estimated in Table D.4 have similar point estimates, albeit with lower statistical significance. For example, with nine free parameters, the regression R^2 is just about 9%. When imposing the IUF restriction with only three free parameters, we already obtain an R^2 about 7% in Table 3. The evidence shows that, besides

[37] Note that the price impact multiplier is the inverse of demand elasticity. See Table 1 of Gabaix and Koijen (2022) for a summary of estimates in the literature.

Table D.3. IV first-stage regression

	factor 1 flow	factor 2 flow	factor 3 flow
concurrent night return	1.20 10 ^{4}	-2.87 10 ^{6}	-5.33 10 ^{6}
	(2.36)	(-1.34)	(-0.91)
lag-1 flow lag-6 flow	0.10	0.16	0.16
	(1.71)	(3.09)	(3.26)
constant	-5.84 10 ^{6}	-0.26 10 ^{6}	0.11 10 ^{6}
regression R^2	4.98%	4.46%	4.28%
F-stats	6.29	5.61	5.37

Note: We report the IV first-stage regression results, where each orthogonalized factor's flow is regressed on its concurrent night return and the difference between one-month and six-month lagged flows. We report in the parenthesis the heteroskedasticity robust t-statistics.

Table D.4. Second-stage regression without IUF restriction

price of risk (regression $R^2 = 8.85\%$)				bootstrapped t-stat			
	MKT	SMB	HML		MKT	SMB	HML
MKT	618	3,147	-3,233	MKT	1.58	1.37	-1.63
SMB	78	3,270	-4,194	SMB	0.17	0.61	-2.21
HML	-48	-340	3,375	HML	-0.35	-0.27	4.32

Note: We report the point estimates and t-statistics of prices of risk without imposing the IUF restriction. Specifically, we directly run the second-stage regression (27) of demeaned returns on factor flows times quantity of risk for the MKT, SMB, and HML factors. We bootstrap 10,000 times to obtain the t-statistics.

the economic contents, our orthogonalization restriction is a tool that increases statistical power while losing little information.

In Table D.5, we present the summary statistics of return correlations between the MPIR strategy and 154 firm characteristics-based anomaly portfolios for the baseline specification (the first row) and several alternative specifications. We see low return correlations between the MPIR strategy and anomaly portfolios.

Table D.5. Correlations between the MPIR strategy and anomaly portfolio returns

	# of obs.	mean	std.	P5	P10	P25	median	P75	P90	P95
with shrinkage	154	-0.04	0.17	-0.34	-0.27	-0.14	-0.04	0.10	0.21	0.23
without shrinkage	154	-0.04	0.17	-0.33	-0.29	-0.14	-0.05	0.08	0.19	0.21
with shrinkage, day return	154	-0.04	0.18	-0.34	-0.29	-0.16	-0.04	0.07	0.21	0.26
without shrinkage, day return	154	-0.03	0.14	-0.30	-0.24	-0.12	-0.03	0.08	0.14	0.19

Note: We present the summary statistics of return correlations between the MPIR strategy and 154 anomaly portfolios. The first row considers the baseline MPIR strategy that applies the shrinkage and uses the price of risk estimated from total returns. The next three rows consider the robustness of not applying the shrinkage or estimating the price of risk using day (open-to-close) returns.