Rethinking Volume*

Philippe van der Beck†, Lorenzo Bretscher‡, Zhiyu Julie Fu§

June 13, 2025

[Click here for the latest version]

Abstract

Gross trading volumes in financial markets are large and far exceed return volatility. In contrast, "net volume" – trading from persistent portfolio reallocations – is substantially lower, as it excludes transitory round-trip trades. This observation reveals a fundamental tension: If return volatility is high, while net volume is low, then market participants either agree with each other (they are "homogeneous"), or they are not sensitive to price changes (they are "inelastic"), resulting in large price impacts of demand shocks. We formalize this tradeoff and demonstrate that the ratio of return volatility to net volume provides a lower bound on price impact, conditional on the level of investor heterogeneity. Using several measures from survey data, we document substantial heterogeneity, implying meaningful lower bounds on price impacts. The bounds align closely with reduced-form estimates from a variety of quasi-experiments, such as price impacts from index reconstitutions, whereas traditional liquidity measures based on gross trading volumes perform poorly. Our bounds prove particularly useful in settings where event-study evidence is difficult to obtain: we demonstrate how they vary over time, across individual assets, and at various levels of aggregation, including the aggregate stock market, and discuss their implications for asset pricing models and the macro-structure of financial markets. We argue in such markets with heterogeneous and inelastic investors, observed trading volumes are not peripheral but central to understanding asset price movements.

*We thank Avinash Sattiraju for excellent research assistance. We thank Malcolm Baker, Xavier Gabaix, Robin Greenwood, Sam Hanson, Ralph Koijen, Erik Stafford, Adi Sunderam, Luis Viceira, and seminar participants at the Demand in Asset Markets Workshop, Harvard Business School, the University of Hong Kong (HKU), and the Hong Kong University of Science and Technology (HKUST) for helpful comments and discussions.

†pvanderbeck@hbs.edu, Harvard Business School

‡lorenzo.bretscher@unil.ch, University of Lausanne, Swiss Finance Institute & CEPR

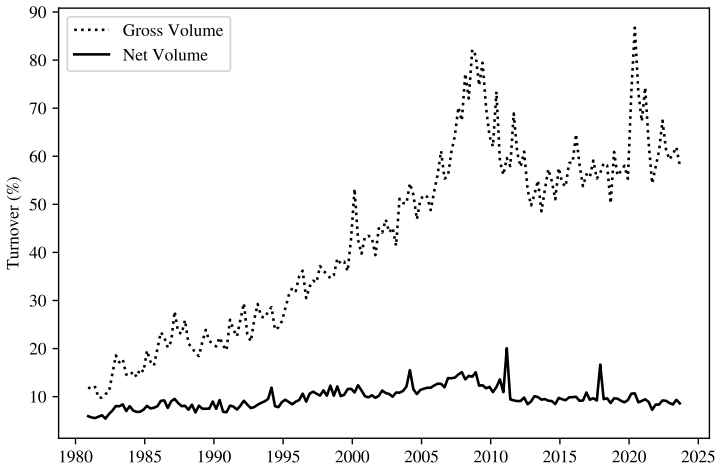
§z.fu@wustl.edu, Washington University in St. Louis, Olin School of Business

1 Introduction

It is widely held that trading activity in financial markets is extraordinarily large: Based on gross (total) trading volume, the entire market value of a typical firm changes hands twice annually. To interpret volume, the literature offers two complementary perspectives. First, investors trade only when they have different views or preferences, so high trading volume has been interpreted as evidence of considerable investor disagreement (e.g., Hong and Stein, 2007). Second, counterparties would take opposite positions only at favorable price adjustments, so high volume relative to return volatility suggests that investors are responsive to price changes, and therefore trades have small price impacts (e.g., Amihud, 2002). Through the lens of these interpretations, the large trading volumes relative to price volatility suggest that investors often disagree, yet their disagreements typically have only minor impacts on prices.

However, gross trading volume can be highly misleading: By including transitory round-trip trades within a period, it significantly overstates actual *net* trading across investors that persists over the period of interest. To better reflect net trading activity, we propose "*net volume*" – the total net portfolio reallocations across all investors over the period of interest, excluding within-period round-trip trades. Figure 1 reveals a stark contrast between gross and net volume: At the quarterly frequency, while gross volume often exceeds 50% of total outstanding shares, net volume is substantially smaller – less than 10% for the average stock. Moreover, this gap has widened dramatically from a factor of two to approximately ten from 1980 to 2025.

Figure 1: **Gross Trading Volume versus Net Volume**
The figure plots total CRSP trading volume (relative to shares outstanding) aggregated at the quarterly frequency against net volume: the total net portfolio changes from quarterly institutional portfolio changes. The sample period is from 1980 to 2024.



The low net trading volumes reveal a fundamental tension between the two complementary views: When net volume is low while return volatility is high, investors cannot be both *heterogeneous* (disagreeing with each other) and *price-elastic* (highly responsive to price changes) at the same time. If investors disagree with each other, they must be unwilling or unable to take opposite positions without large price adjustments, resulting in large price movements with low volumes. Or, if they are sensitive to price, they must share similar beliefs or preferences, so there is limited trading activity among investors, and prices adjust to reflect common beliefs. This fundamental tension between heterogeneity and elasticity forms the core insight of our paper.

We formalize this trade-off in a general framework that connects two observable moments, trading volume σ_q, return volatility σ_p, to the underlying market structure: investor homogeneity ρ (the share of demand shocks explained by the cross-investor average)[1] and the *price impact* $\mathcal{M}$ per 1% demand shock (also referred to as "multiplier").

Our main theoretical result establishes that the price impact $\mathcal{M}$ is bounded below by:

$$\mathcal{M} \geq \frac{\sigma_p}{\sigma_q} \times \sqrt{\frac{1}{\rho} - 1}.$$

The bound captures a key insight: high return volatility relative to volume implies either high price impact (high $\mathcal{M}$) or homogeneous investors (high ρ). To understand this relationship intuitively, consider total demand as an iceberg. Only the heterogeneous component of demand surfaces and becomes visible as observable trading volumes, while the common demand shifts that move all investors in the same direction remain submerged and unobserved. Investor homogeneity ρ determines the relative size of the observed volumes versus the unobserved common demand shocks. When investors are perfectly homogeneous ($\rho \to 1$), observed volumes represent only the tip of the iceberg. The high return volatility is driven by the large unobserved common demand shocks without requiring a large price impact. Conversely, when investors are highly heterogeneous (small ρ), most demand shocks surfaces as observable volumes. Hence, low observed volumes imply small aggregate demand shocks, and large return volatility can only be reconciled with a substantial price impact.

These two scenarios, while both consistent with observed volumes and volatilities, have funda-

[1] Conceptually, ρ is close to the average correlation of demand shocks across investors. We use "demand shocks" to refer to any innovation that leads to a change in investors' portfolio choice problem given the current price. When demand shocks originate solely from beliefs, ρ captures investor *agreement*. More generally, demand shifts may arise from heterogeneous constraints, regulatory frictions, or non-pecuniary preferences. We therefore use the broader term *investor homogeneity* to encompass all forms of variation in investors' portfolio choices.

mentally different implications for how we understand asset prices. With homogeneous and elastic investors, trading volumes are merely a side show – they capture only minor deviations from common demand shifts that drive prices. In this world, asset pricing can rely on representative-agent models while treating trading volume as an irrelevant byproduct of price formation. In a world with heterogeneous and inelastic investors, however, observed trading carries substantial incremental information about asset prices. Here, observed trades are not peripheral but central to understanding asset price movements.

The derivation of our bound relies on a set of fairly general assumptions. Most notably, following the long tradition of log-linearization in finance and the growing literature on demand-system asset pricing (Campbell and Viceira, 2002; Gabaix and Koijen, 2021), we assume that portfolio choice problems can be approximated to first order by linear demand curves. Importantly, we impose no structural assumptions on the source or structure of demand shocks, nor on particular microfoundations of the elasticities. This makes our bound *empirical* in nature, similar in spirit to the Hansen–Jagannathan bound (Hansen and Jagannathan, 1991). Due to the model-agnostic nature of the bound, it can serve as a diagnostic tool when developing structural models to rationalize these moments. It can also be used by empirical studies estimating price impact $\mathcal{M}$ or heterogeneity ρ to back out the other parameter, thus providing a more comprehensive picture of the market environment.

We apply our bound to individual U.S. stocks. While net volume and return volatility are directly observable, investor homogeneity ρ is inherently unobserved. However, there is substantial empirical evidence confirming that investors are highly heterogeneous ($\rho \ll 1$). Investors differ markedly in regulatory constraints, return expectations, trading patterns, and portfolio compositions.[2] Given this evidence, a highly elastic market at the quarterly frequency appears unlikely through the lens of our bound: For example, as $\frac{\sigma_p}{\sigma_q} \approx 1$ for the average stock, achieving a price impact below 0.1 requires almost perfect homogeneity among investors, i.e., $\rho > 0.99$.

To estimate the price-impact bound, we incorporate empirical proxies for ρ. Our goal is not to obtain precise point estimates, but to demonstrate that the homogeneity parameter ρ lies within a moderate range—avoiding pathological extremes near zero or one. Our baseline proxy measures homogeneity through the common variations in forecast updates on earnings across analysts. For the

[2]See, among others, Giglio et al. (2021), Koijen and Yogo (2019), Dahlquist and Ibert (2024), Couts et al. (2024), Barber and Odean (2008), Guiso et al. (2008), Kandel and Pearson (1995), Barber and Odean (2001), and Bretscher et al. (2025). In addition, the vast investor heterogeneity can be directly observed by the fact that the number of different mutual funds catering to the preferences and beliefs of different investors exceeds the number of stocks in the U.S. (see Investment Company Institute Investment Company Institute (2025)).

average U.S. firm, forecast updates in earnings per share (EPS) across analysts explain approximately 58% of the total variation in one-quarter-ahead EPS updates, yielding $\rho = 0.58$. When we use long-term growth (LTG) forecasts instead, we obtain $\rho = 0.27$. Applying these homogeneity measures to our model generates stock-level price impacts of 0.75 and 1.0, respectively.

The model parameter ρ should reflect all forms of investor heterogeneity, including belief disagreement, constraints, and preferences, which may not be fully reflected in our empirical proxies for ρ. However, our measures suggest that the true value of ρ lies in the moderate range rather than at the extremes. In the moderate range, $\sqrt{\frac{1}{\rho} - 1}$ is relatively flat, making it insensitive to variations in ρ. Consequently, cross-stock variation in our bounds is driven primarily by the volatility-to-volume ratio $\frac{\sigma_p}{\sigma_q}$. In fact, a simplified multiplier estimate $\tilde{\mathcal{M}} \equiv \frac{\sigma_p}{\sigma_q}$ (assuming $\rho = 0.5$) performs nearly as well as the general bound.

We validate our price impact bounds against well-documented demand shock events, including S&P 500 index inclusions and mutual fund flow-induced trading. Our bound-implied price impact estimates exhibit strong correlations with actual price movements across these events. Stocks with larger bounds experience significantly higher price changes for given demand shocks. For instance, the price impact of flow-induced trades increases monotonically with our bounds. In contrast, traditional liquidity measures based on gross trading volume, such as Amihud's (2002) illiquidity ratio, show no significant explanatory power for price impacts of persistent demand shifts. Constructing our bounds with gross (rather than net) volume reveals no significant relationship to event study price impacts, highlighting that net volume is not merely a modeling choice but a meaningful economic quantity.

We further examine how price impact bounds vary over time, across assets, and at different aggregation levels. First, long-term price impact has remained largely unchanged over the past 30 years despite gross trading volume increasing fivefold. As Figure 1 shows, net volume has remained relatively constant. Second, consistent with information-based theories, large-cap stocks exhibit *smaller* price impacts, while stocks with higher systematic risk show *larger* price impacts, consistent with risk-based foundations. Momentum stocks also display higher price impacts, reflecting upward-sloping demand curves of momentum investors. Third, examining different aggregation levels reveals that net volumes decline more rapidly than return volatility when moving to higher aggregation levels, resulting in higher price impact estimates for market-level portfolios compared to individual stocks or industry portfolios.

Our bounds provide an ex-ante measure of how informative trading volume is for security prices. When demand is highly elastic, trading volume reveals little about prices since demand shifts are

easily accommodated with minimal price impact. Conversely, when demand is inelastic, trading volume becomes highly informative for prices. Empirically, our bounds suggest that the stock market is closer to inelastic and heterogeneous than elastic and homogeneous. In such markets, observed trading volumes capture a significant portion of the underlying demand variation, and have significant impact on prices. Echoing Hong and Stein (2007) and more recently Gabaix and Koijen (2021), our bounds demonstrate that trading volumes are not mere byproducts of price formation, but essential for understanding asset prices, and financial market volatility.

Related Literature. This paper offers a synthesis of the literature on price impact and the literature on investor disagreement. First, a large strand of literature documents that investor-specific demand shocks can have a meaningful long-term price impact. For example, a series of papers studies price changes upon inclusion or deletion of a stock in an index (see, among others, Shleifer, 1986, Harris and Gurel (1986), Wurgler and Zhuravskaya (2002), Kaul et al. (2000), Chang et al. (2015), Pavlova and Sikorskaya (2022), Greenwood and Sammon (2025), and Aghaee (2025)). Coval and Stafford (2007), Lou (2012), and Edmans et al. (2012) document persistent price changes due to flow-induced trading by mutual funds. Hartzmark and Solomon (2021), Schmickler and Tremacoldi-Rossi (2022), Kvamvold and Lindset (2018), and Honkanen et al. (2025) document that reinvested dividend payouts significantly affect asset prices. Our price impact bounds provide an ex ante statistic that can be used as a simple benchmark based on observable empirical moments. In addition, finding exogenous variation in demand to identify price impact for aggregated portfolios such as the total stock market is often difficult. Our bounds can easily be computed at different levels of aggregation for all asset classes and, thus, provide a useful sanity check for event studies. More importantly, event-study evidence is often difficult to obtain, particularly when researchers face limited cross-sectional variation over time or insufficient time-series variation across assets. Estimating price impacts for aggregated portfolios – such as entire asset classes – is especially challenging, as it requires identifying demand shifts that are both exogenous and sufficiently large. Our estimation-free bounds offer a practical alternative by providing theoretically grounded benchmarks for the expected price impact in settings where event studies are infeasible.

More broadly, our bounds contribute to the burgeoning demand-system asset-pricing literature that models investors' portfolio allocation and asset prices jointly (Koijen and Yogo, 2019; Gabaix and Koijen, 2021).[3] Several recent papers in this literature have alluded to the tension between

[3]See, among others, Koijen and Yogo (2020), Haddad et al. (2021), Han et al. (2021), Koijen et al. (2021), Fang et al.

elasticity and heterogeneity. Gabaix and Koijen (2021) argue that the relatively stable equity share of institutional investors implies inelasticity at the aggregate market level. We share a similar model framework with their approach. Their granular-instrument-variable (GIV) estimator further imposes a factor structure on demand shocks and identifies investor homogeneity by extracting common factors from investor flows. More recently, Gabaix et al. (2025) compute risk transfer – changes in market risk exposure by households, a measure conceptually close to the net volume for the aggregate market – is very small at the quarterly frequency. They demonstrate that standard macro-finance models featuring high price elasticities cannot reconcile the tension between the heterogeneity in holdings and the small risk transfer in flows. Complementary to their approach, our bounds take a relatively *model-free* approach. Moreover, we compute bounds for individual stocks, different portfolios, and the aggregate stock market, and test their empirical relevance in event studies.

Our paper also contributes to the literature on investor disagreement. For example, Kandel and Pearson (1995) and Bamber et al. (1999) document that earnings announcement days consistently feature abnormal trading volume and small price changes. In those papers, the combination of high volume and low volatility is typically interpreted as evidence of differential interpretations of public signals, i.e., *disagreement*. Hong and Stein (2007) advocate for models featuring disagreement given the enormous trading volume observed even at times when return volatility is low.[4] Our paper offers a formal framework to interpret volumes as investor disagreement. While we find that net trading volumes are low at longer horizons, we argue this does not reflect an absence of disagreement. Instead, it reflects the inelasticity of market participants, which amplifies the price impact of investor-specific demand shocks and serves as an empirically useful measure of long-term price impact.

Third, by drawing a distinction between net and gross volume, our paper naturally contributes to the literature on market liquidity.[5] We highlight that the distinction between net and gross volume helps explain why traditional liquidity measures – such as those in Amihud (2002), Pástor and Stambaugh (2003), and Brennan et al. (2013) – may be less suitable to capture long-term price impact from persistent demand shifts. Intuitively, our bounds on price impact can be viewed as a long-horizon

(2022), Coqueret (2022), Huebner (2023), Jiang et al. (2022), Jiang et al. (2024), Koijen et al. (2024), Jansen (2025), Tamoni et al. (2024), Bretscher et al. (2025), Chaudhary et al. (2024), Jansen et al. (2024), Darmouni et al. (2022).

[4]See, for example, Harris and Raviv (1993) and Banerjee and Kremer (2010) for theoretical models reconciling the observed empirical patterns.

[5]See, among others, Constantinides (1986), Brennan and Subrahmanyam (1996), Heaton and Lucas (1996), Vayanos (1998), Brennan et al. (1998), Datar et al. (1998), Chordia et al. (2001), Amihud (2002), Jones (2002), Huang (2003), Pástor and Stambaugh (2003), Anshuman and Viswanathan (2005), Brunnermeier and Pedersen (2009), Garleanu and Pedersen (2007), and Bouchaud (2022).

counterpart to the illiquidity measure proposed by Amihud (2002).

The remainder of the paper is structured as follows. Section 2 lays out the main theory. Section 3 describes the data, construction of net volume and discusses its difference from gross trading volume. Section 4 constructs the price impact bounds for the cross-section of US equities; Section 5 tests the empirical relevance of the bounds using different event studies. Motivated by the empirical relevance, Section 6 then examines the heterogeneity of these bounds outside of event studies – over time, across assets, and at different levels of aggregation. Section 7 concludes.

2 Theory

In this section, we lay out our main framework and derive the price impact bound.

Notation. Throughout, we use $i = 1, 2, ..., I$ to denote the investor, n to denote the asset. We use $S_i(n)$ to denote the ownership share of investor i in the market for asset n. As a short-hand, we use subscript S in place of i to denote size-weighted aggregation, i.e., $x_S(n) = \sum_{i=1}^{I} S_i(n)x_i(n)$. To highlight the cross-sectional expectation-like feature of the size-weighted aggregation, we also use $\hat{\mathbb{E}}_S^{cs}[x_i] = \sum_{i=1}^{I} S_i x_i$, and suppress the subscript S when there is no ambiguity.

2.1 The Demand Curve

Consider a generic portfolio allocation $Q_{i,t}(n) = Q_i(P_t(n), U_{i,t})$, where $Q_{i,t}(n)$ is the quantity of asset n held by investor i at time t, $P_t(n)$ is the price of asset n at time t, and $U_{i,t}$ captures all other factors that affect investor i's demand for asset n at the given price $P_t(n)$. These factors can include the investor's wealth, the risk-free rate, risk aversion, uncertainty, prices of substitutable assets, and other relevant variables.

We take a log-linear approximation of the portfolio choice problem around the long-run mean and take first-differences to obtain a linear demand curve:

$$\Delta q_{i,t}(n) = -\zeta_i(n)\Delta p_t(n) + u_{i,t}(n) \tag{1}$$

where $\Delta q_{i,t}(n)$ is the percentage change in holdings of asset n by investor i at time t (which we refer to as *flows*), $\Delta p_t(n)$ is the percentage price change of asset n at time t, or simply referred to as its return at time t, and $u_{i,t}(n)$ is the demand shock for investor i at time t. For simplicity, we assume their

time-series mean is equal to zero. The parameter $\zeta_i(n)$ is the investor-asset-specific elasticity, which measures how much investor i's demand for asset n changes when the price changes by 1%, *ceteris paribus*.

To provide intuition for the different components of the demand curve, we can connect this linear specification with canonical models. In Appendix B, we sketch several microfoundations, including standard portfolio choice under CRRA utility. Our preferred interpretation draws on learning-from-price models such as Grossman and Stiglitz (1980) and Hellwig (1980). In these models, the demand shock $U_{i,t}$ represents noisy private signals about the asset's fundamental value, while the price $P_t(n)$ aggregates information across investors. The price elasticity $\zeta_i(n)$ captures the trade-off between the informativeness of one's private signal and the market price: the more accurate the market price relative to the investor's private signal, the more the investor relies on the price, resulting in a more inelastic demand curve (smaller $\zeta_i(n)$).

While learning-from-price models provide a natural framework for interpreting elasticity and disagreement, we do not restrict our analysis to this interpretation. Instead, we specify the demand curve generically. In any asset-pricing model that features portfolio choice, either explicitly or implicitly, investor demand can be decomposed into changes due to price movements and changes holding prices fixed.[6] Our bound holds under these different model frameworks; however, one should be careful, when interpreting the implied elasticities and demand shocks. Moreover, the underlying model does not need to be static: in dynamic settings, investors care about the path of future expected returns, while the market clears through the current price, which summarizes the market's expectations about future returns. In this case, investors' demand shocks contain beliefs about future expected returns that deviate from those implied by the current price. Further, our framework can also accommodate multiple assets. To that end, the demand shock $u_{i,t}(n)$ captures substitution and arbitrage effects across assets.[7] Our only assumption at this stage is that first-order log-linearization provides a reasonable approximation of the true portfolio choice problem.

For a generic demand curve specified in Equation (1), our goal is to connect elasticity and demand

[6] See Koijen and Yogo (2025) for further discussion on microfoundations.

[7] With multiple assets, the full demand curve can be specified in vector form as $\Delta \mathbf{q}_{i,t} = -\mathbf{Z}_i \Delta \mathbf{p}_t + \tilde{\mathbf{u}}_{i,t}$. For asset n, this can be expressed as $\Delta q_{i,t}(n) = -\zeta_i(n)\Delta p_t + \tilde{u}_{i,t}(n)$, where $u_{i,t}(n) \equiv \sum_{m \neq n} Z_{i,(n,m)} \Delta p_t(m) + \tilde{u}_{i,t}(n)$. Care needs to be taken when interpreting demand shock homogeneity, as it includes substitution effects; the inverse of elasticity in this case also corresponds to the price impact *relative to* its substitute assets. When substitution effect is mild, as it is the case for the equity market (e.g., see Chaudhary et al., 2023), the difference is quantitatively small. However, this model framework would not be suitable for analyzing markets with strong substitution patterns such as individual corporate bonds.

shock heterogeneity to observable moments: volatilities in returns and flows. To do so, we first study how the return and flow volatilities are determined in the log-linear model.

In the remainder of this section, we proceed stock by stock, and suppress the stock index n for notational ease.

2.2 Elasticity and Price Impact

The flip side of the elasticity is the price impact per unit of demand shock. To see that, we impose the market clearing condition – all trades sum to zero. Denote $S_i = \frac{Q_i}{\sum_i Q_i}$ the ownership share of investor i in the market for asset n. The market clearing condition is given by:

$$\sum_i S_i \Delta q_{i,t} = 0 \tag{2}$$

Price adjusts to clear the market, and hence,

$$\Delta p_t = \frac{1}{\zeta_S} u_{S,t} \tag{3}$$

where $\zeta_S = \sum_i S_i \zeta_i$ and $u_{S,t} = \sum_i S_i u_{i,t}$ are the aggregate elasticity and demand shift given by the size-weighted averages of investor-specific elasticities and investor-specific demand shifts respectively. The inverse of the aggregate elasticity, $\frac{1}{\zeta_S}$, quantifies how much the price adjusts for a 1% aggregate demand shock of total outstanding shares. Therefore, the lower the aggregate demand elasticity, the larger is the price adjustment per unit of demand shock which is needed to induce investors to step in. We denote the inverse of the aggregate demand elasticity as $\mathcal{M}$ and refer to it as *price impact* or *multiplier* of asset n.

The multiplier $\mathcal{M}$ links return volatility to the volatility of the aggregate demand shock, given by:

$$\sigma_p^2 = \mathcal{M}^2 Var(u_{S,t}) \tag{4}$$

For example, through the lens of this framework, the well-known excess volatility puzzle implies that either standard models do not generate sufficiently volatile aggregate demand shocks, or that agents are too responsive to price changes, i.e., the price impact $\mathcal{M}$ is too small.

2.3 Flows and Heterogeneity

To illustrate the relationship between flows and heterogeneity, we first consider the case with homogeneous elasticities across investors: $\zeta_i = \zeta_S = \zeta$. Note that this assumption will be relaxed later. Under the homogeneous elasticity assumption, we can plug the equilibrium price equation (3) into the demand equation (1) to have:

$$\Delta q_{i,t} = u_{i,t} - u_{S,t}. \tag{5}$$

Hence, trades reflect the *differences* of investors' demand shocks from the average demand shock in the market.

The size-weighted average variance of $\Delta q_{i,t}$ is given by:

$$\begin{aligned} \sigma_q^2 &\equiv \sum_{i=1}^{I} S_i Var\left(\Delta q_{i,t}\right) \\ &= \left(\sum_{i=1}^{I} S_i Var(u_{i,t})\right) - Var\left(u_{S,t}\right) \end{aligned} \tag{6}$$

To derive the second equality, we use Equation (5) and the identity $\sum_{i=1}^{I} S_i Cov(u_{i,t}, u_{S,t}) = Var(u_{S,t})$. Hereafter, we refer to σ_q as *flow volatility*. It measures the total amount of trading activity by investors. The theoretical analysis focuses on flow volatility defined in (6); later we show that (net) trading volume is a close proxy for flow volatility, and use the terminology interchangeably when the distinction is unimportant.

Defining $\rho \equiv \frac{Var(\sum_{i=1}^{I} S_i u_{i,t})}{\sum_{i=1}^{I} S_i Var(u_{i,t})}$, we can rewrite flow volatility as follows:

$$\sigma_q^2 = Var(u_{S,t}) \left(\frac{1}{\rho} - 1\right). \tag{7}$$

We refer to ρ as *investor homogeneity*. To understand the interpretation, note that it is the share of demand shocks that is explained by the cross-sectional average of the demand shocks. To see this most clearly, we can use the cross-sectional expectation notation $\hat{\mathbb{E}}^{cs}$ to express it as follows:

$$\rho = \frac{Var\left(\hat{\mathbb{E}}^{cs}\left[u_{i,t} \mid t\right]\right)}{\hat{\mathbb{E}}^{cs}\left[Var\left(u_{i,t} \mid i\right)\right]} \in [0, 1]. \tag{8}$$

Empirically, ρ is the R^2 of the (size-weighted) time fixed effects of the demand shocks. As an R^2, it ranges between 0 and 1. When $\rho = 1$, all investors have identical demand shocks, and hence

are homogeneous; when $\rho \to 0$, the demand shocks are completely heterogeneous across investors.[8] Alternatively, with homoskedasticity, the homogeneity ρ can also be interpreted as the average pairwise correlation of the demand shocks $\rho \approx \sum_{i=1}^{I} \sum_{j \neq i} S_i S_j corr(u_{i,t}, u_{j,t})$.[9]

Note that heterogeneity does not only come from differences in idiosyncratic demand shocks, but also from differences in the responses of different investors to the same shocks. To see this, suppose idiosyncratic demand shocks are determined by their differential exposure, λ_i, to a single common shock, η_t, which has unitary variance, i.e., $u_{i,t} = \lambda_i \eta_t$. Use $\hat{\mathbb{E}}^{cs}$ to denote the size-weighted cross-sectional average, we have:

$$\rho = \frac{\hat{\mathbb{E}}^{cs}\left[\lambda_i\right]^2}{\hat{\mathbb{E}}^{cs}\left[\lambda_i^2\right]} = \left(1 + \frac{\widehat{Var}^{cs}\left(\lambda_i\right)}{\hat{\mathbb{E}}^{cs}\left[\lambda_i\right]^2}\right)^{-1}. \tag{9}$$

This implies that demand homogeneity decreases in the variation of the exposures to the common shock η_t across investors. Further, demand homogeneity can be arbitrarily close to zero when the variation in λ_i relative to its mean is large. In sum, investors can be highly heterogeneous even if their idiosyncratic demand shocks can be fully explained by a common shock, η_t, provided their exposures to that shock differ.

With this interpretation, Equation (7) states that flow volatility is the product of the size of aggregate demand shocks and the investor heterogeneity. The extent to which highly volatile aggregate demand translates into flow volatility (trading volume) is driven by the amount of investor disagreement ρ.

2.4 The Elasticity Bounds

To derive the bound, the key observation is that both price and flow volatilities depend on the volatility of the average demand shock in the market, but with a different coefficient: the multiplier $\mathcal{M}$ for return volatility and the investor heterogeneity $\mathcal{D}$ for flow volatility. Taking the ratio of flow volatility

[8] With finite number of investors, $\rho \geq \frac{\sum_i S_i^2 \sigma_i^2}{\sum_i S_i \sigma_i^2}$ if the covariances of the demand shocks across investors are non-negative, and it reaches the lower bound when shocks are completely uncorrelated. It reaches zero only if investors demand completely offset each other in aggregate.

[9] We can write $Var(u_{S,t}) = \sum_{i=1}^{I} S_i^2 \sigma_i^2 + \sum_{i=1}^{I} \sum_{j \neq i} S_i S_j \sigma_i \sigma_j corr(u_{i,t}, u_{j,t})$, under homoskedasticity, $\sigma_i = \sigma_j = \sigma$, so we have $\rho = \sum_{i=1}^{I} S_i^2 + \sum_{i=1}^{I} \sum_{j \neq i} S_i S_j corr(u_{i,t}, u_{j,t})$. The first term is the Herfindahl–Hirschman Index (HHI) of the ownership distribution, which vanishes to zero as N is large.

in Equation (7) and return volatility in Equation (4), we have:

$$\mathcal{M} = \frac{\sigma_p}{\sigma_q} \times \underbrace{\sqrt{\frac{1}{\rho} - 1}}_{\mathcal{D}} \tag{10}$$

We refer to $\mathcal{D}$ as *investor heterogeneity*.

Equation (10) connects observable market quantities – price and flow volatilities – to the underlying elasticity and investor heterogeneity. When prices exhibit high volatility relative to trading activity (a large $\frac{\sigma_p}{\sigma_q}$ ratio), two possible explanations emerge: either the price multiplier $\mathcal{M}$ is large, amplifying price responses to demand shocks, or investors are highly homogeneous ($\rho \to 1$), causing observed trading activity (the tip of the iceberg of total demand shocks) to significantly under-represent the magnitude of underlying demand shocks.

So far, price impact $\mathcal{M}$ is derived under the homogeneous-elasticity assumption. To consider the case with heterogeneous elasticities, we make an assumption on the distribution of elasticities. Without getting too attached to a particular data-generating process, we consider the following environment:

Assumption 1. *The elasticity ζ_i for each investor i are drawn independently from the parameters governing the demand shock process $u_{i,t}$.*

Assumption 1 serves as a neutral benchmark, but it is not necessary for the main result. With an arbitrary data-generating process of elasticities and the demand shocks, one can end up in the pathological case where the investors that receive larger demand shocks end up selling because they also react more to the price changes. In Appendix A, we discuss the more precise condition under which our main theorem holds.

Under the Assumption 1, the implied multiplier will be even larger for a given level of investor heterogeneity $\mathcal{D}$. Intuitively, this is because heterogeneous elasticities provide additional reasons for trading other than investor heterogeneity: Consider the case where investors receive an identical demand shock. Price adjusts, but as investors respond to the price adjustment differently, they also want to trade with each other.

Hence, when an econometrician infers the magnitude of aggregate demand shocks from observed flow volatility under the assumption of homogeneous elasticities, this assumption leads to *overestimation* of the underlying demand shocks. The overestimation occurs because some observed trading activity stems not from heterogeneous demand shocks but from investors' heterogeneous responses to

price changes. Since the true aggregate demand shocks are smaller under heterogeneous elasticities, the actual price impact exceeds that given by (10). Formally, we establish the following theorem:

Theorem 1. *Under Assumption 1, the price impact $\mathcal{M}$ of demand shocks is lower-bounded by the volatility-to-volume ratio $\frac{\sigma_p}{\sigma_q}$, adjusted by the correlation of investors' demand shifts ρ:*

$$\mathcal{M} \geq \frac{\sigma_p}{\sigma_q} \times \sqrt{\frac{1}{\rho} - 1} \tag{11}$$

Proof. See Appendix A. □

In-sample bounds. Notice that though we express the bound in terms of population parameters, the identities used in deriving the bound all hold in sample as well. Hence we can express the bound using sample moments, given as:

$$\mathcal{M} \geq \frac{\hat{\sigma}_p}{\hat{\sigma}_q} \times \sqrt{\frac{1}{\hat{\rho}} - 1} \tag{12}$$

where $\hat{\sigma}_p$ and $\hat{\sigma}_q$ are the sample counterparts of price and flow volatilities, and $\hat{\rho}$ is the homogeneity of the demand shocks within the sample period.

In fact, the bound can be applied period by period, under the assumption that $\Delta q_{i,t}$ and Δp_t have mean zero in a given period t (which can be achieved by demeaning across t, assuming means are stable):

$$\mathcal{M}_t \geq \frac{|\Delta p_t|}{\sqrt{\sum_{i=1}^{I} S_i \Delta q_{i,t}^2}} \times \sqrt{\frac{1}{\rho_t} - 1} \tag{13}$$

where $\rho_t \equiv \frac{(\sum_{i=1}^{I} S_i u_{i,t})^2}{\sum_{i=1}^{I} S_i u_{i,t}^2}$ is the investor homogeneity *in period t*.

2.5 Flow Volatility and Net Volume

The key input to our bound, flow volatility σ_q, is defined as the size-weighted average of investor-specific flow volatilities. Seemingly complicated, we show that it is closely related to the total trading activity from changes in investors' portfolios, which we term *net volume*. For a stock in a given quarter t, net volume is defined as the sum of absolute value of quarter-on-quarter changes in positions of all investors, normalized by shares outstanding:

$$\text{NetVol}_t = \frac{\sum_i |\Delta Q_{i,t}|}{\bar{Q}} \tag{14}$$

where $\Delta Q_{i,t} = Q_{i,t} - Q_{i,t-1}$ is the change in position of investor i from $t-1$ to t, and $\bar{Q}$ is total supply.

The net volume measures the (size-weighted) *mean absolute deviation* (MAD) of flows:

$$\mathbb{E}\big[\text{NetVol}_t\big] = \mathbb{E}\left[\sum_i S_i \frac{|\Delta Q_{i,t}|}{S_i \bar{Q}}\right] = \sum_i S_i \mathbb{E}\left[|\Delta q_{i,t}|\right]. \tag{15}$$

It mirrors the definition of flow volatility, $\sigma_q \equiv \sqrt{\sum_{i=1}^{I} S_i \mathbb{E}\left[(\Delta q_{i,t})^2\right]}$, but with an $\mathcal{L}_1$-norm rather than an $\mathcal{L}_2$-norm. For common distributions, the mean absolute deviation $\mathbb{E}\left[|\Delta q_{i,t}|\right]$ is proportional to standard deviation $\sigma_{q,i}$ by a constant factor ν determined by the underlying distribution. For example, for normally distributed $\Delta q_{i,t}$, $\nu = \sqrt{\frac{\pi}{2}} \approx 1.25$. Empirically, Appendix Figure E.1 shows that net volume scaled by $\sqrt{\frac{\pi}{2}}$ and σ_q are effectively equivalent with a cross-sectional correlation around 0.9 and an OLS slope coefficient of 1.1. For this reason, we view the scaled net volume, $\sqrt{\frac{\pi}{2}}\text{NetVol}_t$, as an alternative (and more robust) estimator for σ_q, and use net volume to refer to σ_q at the conceptual level.

Net volume is closely linked to the gross trading volume by construction. However, unlike gross trading volume, which aggregates *all* trades within a quarter, net volume omits offsetting round-trip trades and measures net quarter-on-quarter change in portfolio holdings. To see this, consider an investor that moves from 1000 shares at t to 1100 shares at $t + \frac{1}{2}$, back to 1000 shares at $t + 1$. While the investor's gross volume is 200 shares, their net volume from t to $t + 1$ is $|\Delta Q_{i,t}| = 0$ shares. To understand the liquidity provision at the quarterly frequency (here t to $t + 1$), the intraquarter round trips are irrelevant and hence netted out from net volume. The next section provides more details on the distinction between net volume and gross trading volume.

3 Data and Empirical Facts

3.1 Data Sources and Variable Construction

Data. Our empirical analyses are all at the quarterly frequency. We obtain quarterly institution-level share holdings $Q_{i,t}(n)$ from 1990 to 2024 from Thomson Institutional Holdings Database (s34 file). Institutions are denoted by $i = 1, ..., I$. The subscript t indicates the report date of the 13F filing. [10] Further details can be found in Appendix C.1. Subsequently, we merge quarterly stock holdings

[10] In the main text, we use holdings at the institution level (e.g., BlackRock as a single entity rather than as individual funds) to achieve the most comprehensive coverage. Since holdings are aggregated across funds within the same asset manager, transactions among funds within the same institution are not observed at this level, which could potentially result in smaller net volumes relative to gross volumes. However, in Appendix C.1, we show that net volumes computed

with data on prices and fundamentals from CRSP, Compustat, and IBES. We restrict our sample to common ordinary shares (share codes 10 and 11) traded on the NYSE, AMEX and NASDAQ (exchange codes 1, 2, and 3), that have (on average) at least 10 institutional holders and at least 30% observed institutional ownership.[11] Δ denotes quarterly changes. Ownership shares (size-weights) are denoted by $S_{i,t}(n) = \frac{Q_{i,t}(n)}{\bar{Q}_t(n)}$, where $\bar{Q}_t(n)$ are the total shares outstanding of the stock. Empirically, we do not observe the holdings of *all* investors, but are restricted by reported 13F filings. We therefore construct the trades of the residual sector that holds the remaining shares outstanding such that the trades of all investors sum to 0.[12]

Estimating volatilities. As discussed in Section 2, our bound holds both in sample as well as period by period. We estimate both $\sigma_q(n)$ and $\sigma_p(n)$ in the time series for each stock using 5-year backward-looking rolling windows which prevents our results from suffering from forward-looking bias. We estimate $\sigma_p(n)$ using the time-series volatility of quarterly stock returns. As described in Section 2.5, $\sigma_q(n)$ can either be measured directly as $\sqrt{\sum_{i=1}^{I} S_{i,t}\widehat{Var}(\Delta q_{i,t}(n))}$, the size-weighted average of investor-specific volatilities (the $\mathcal{L}_2$ norm), or approximated using net volume $\sqrt{\frac{\pi}{2}}\hat{\mathbb{E}}\left[\text{NetVol}_t(n)\right]$ (the $\mathcal{L}_1$ norm). We construct both measures and find similar results. Generally, we favor net volume for several reasons. First, it is straightforwardly constructed and closely linked to gross trading volume. Second and more importantly, $\mathcal{L}_2$ norms, such as the standard deviation, are susceptible to outliers – a common feature in flow data – while $\mathcal{L}_1$ norms, such as mean absolute deviation, are more robust estimators of statistical dispersion in the presence of fat tails (due to frequent extensive margin trades).

Unlike σ_q and σ_p, which are directly observable from trade and price data, investor homogeneity ρ is inherently unobserved. To that end, we first present results that are agnostic about the level of investor homogeneity. Later in Section 4.2, we then present and discuss different strategies of how to empirically measure ρ.

The top panel of Table 1 reports $\sigma_p(n)$ and $\sigma_q(n)$ (both measured via $\mathcal{L}_1$ and $\mathcal{L}_2$ norms). The average share in our sample has a quarterly return volatility $\sigma_p(n)$ of 22%. The average σ_q constructed from net volumes is 25%. The 5th percentile, median, and 95th percentile are given by 8%, 23%, and 50% respectively. In contrast, the $\mathcal{L}_2$ measure of σ_q is distributed very similarly with a slightly higher

at the mutual fund level are very close to those at the institution level, suggesting that within-fund-family trades are negligible.

[11] All results are robust to alternative cut-offs.

[12] All results in the paper are robust to omitting the residual sector and constructing $\bar{Q}_t(n)$ (and the corresponding size weights) as the sum of institutional shares held. However, we prefer to construct the residual sector as doing so effectively accounts for trades by the institutional sector as a whole, which would be omitted otherwise.

average of 30% and the 5th percentile, median, and 95th percentile given by 10%, 30%, and 53%, respectively. As a consequence, the ratio of return volatility to net volume (hereafter, volatility-to-volume ratio) equals 1.15 for the average share. However, there exists considerable variation in this ratio as can be seen from 5th and 95th percentiles which equal 0.36 and 3.02, respectively.

Finally, Table 1 also reports moments on the distributions of institutional ownership and trading volume. For example, the average share is held by about 200 institutions with an average institutional ownership share of 60%. Notably, all our main results hold when restricting the sample to stocks for which the institutional ownership is higher than 90%.

Table 1: **Summary Statistics**
The table summarizes the distribution of the key variable inputs for deriving the price impact bounds over the cross-section of US equities. The first rows report the volatility of returns σ_p and the volatility of flows σ_q both explicitly computed via size-weighted investor-specific volatilities, and the $\mathcal{L}_1$ approximation from scaled net volume. The volatilities are computed over 5-year rolling windows. The middle panel reports the number of investors holding each stock, the distribution of institutional ownership and the investor concentration defined as $\sum_i S_{i,t}^2(n)$. The last rows report gross quarterly trading volume (from CRSP) alongside net volume divided by 2. The division by 2 avoids double-counting trades and ensures comparability to gross trading volume.

	Mean	Std	5th pctl.	Median	95 pctl.
Volatilities of Trade and Returns					
Return Volatility σ_p	0.22	0.15	0.09	0.19	0.47
Flow Volatility σ_q ($\mathcal{L}_2$)	0.30	0.13	0.10	0.30	0.53
Net Volume σ_q ($\mathcal{L}_1$)	0.25	0.13	0.08	0.23	0.50
Volatility-Volume Ratio σ_p/σ_q	1.16	1.32	0.36	0.81	3.02
Ownership Distribution					
Number of Institutional Holders	202.94	276.54	14.00	124.00	663.00
Institutional Ownership	0.60	0.25	0.16	0.62	0.99
Ownership Concentration (HHI)	0.23	0.21	0.04	0.16	0.67
Trading Volume					
CRSP Total Volume	0.46	0.55	0.06	0.31	1.35
$\frac{1}{2}$ Net Volume	0.10	0.09	0.02	0.08	0.24

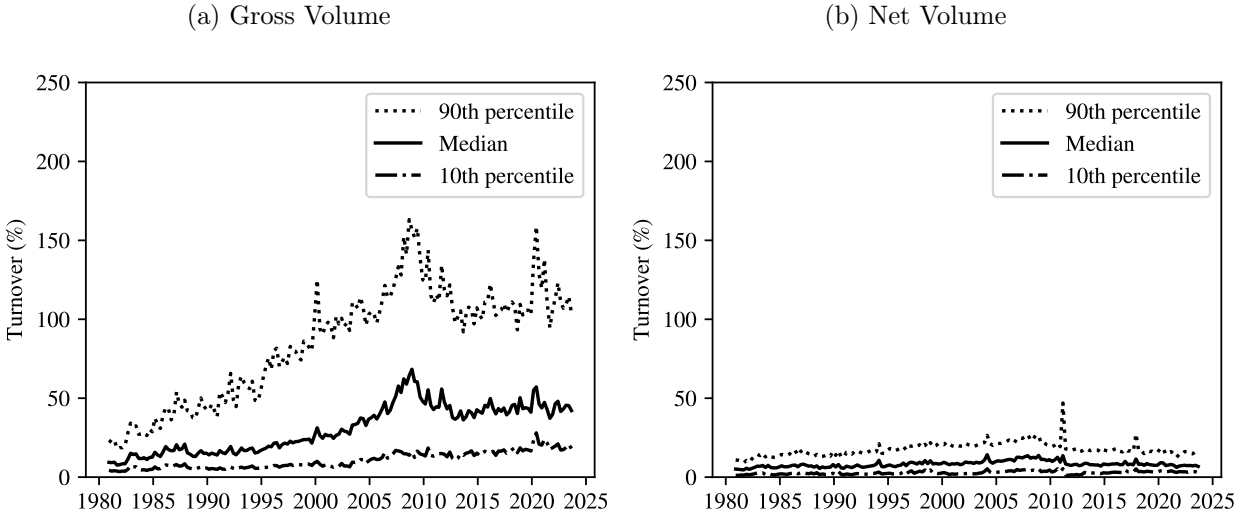
3.2 Net Volume versus Gross Trading Volume

A well-known feature of equity markets is the high volume of trading. In fact, in the past 30 years, the quarterly turnover for the median stock on the NYSE, AMEX, and NASDAQ, was around 50-100%. We confirm this in our sample by computing quarterly gross trading volume by aggregating CRSP monthly turnover at the quarterly frequency. Panel a) of Figure 2 plots the distribution of gross

volume. Over our sample period average quarterly gross trading volume has steadily increased from about 15% to 50%. In contrast, Panel b) of Figure 2 plots net volume constructed from changes in institutional investors' portfolios. The main take away from comparing the two Panels a) and b) is that average gross volume is much higher compared to average net volume. For example, as of 2024, the quarterly net volume amounts to 8% of the shares outstanding for the average stock. In other words, every quarter, institutional investors turn over 8% of a stock's shares outstanding. Instead, quarterly gross trading volume for the average stock is 60%, more than seven times higher than net volume, on average. This ranking is preserved over time and across the entire cross-section of US equities.

If institutional investors were the only investors trading the underlying securities, then any difference between gross and net volume is due to offsetting round-trip trades within a quarter. Importantly, round-trip trades cannot effectively accommodate persistent demand shifts. In other words, this implies that in recent years, 85% of gross trading volume are due to offsetting round-trip trades within a quarter that do not provide quarterly liquidity to financial markets.

Figure 2: **Quarterly Gross Volume versus Net Volume: US Equities**
The figure plots the distribution of quarterly gross trading volume relative to shares outstanding and quarterly net volume divided, over the cross-section of US stocks from 1980 to 2024. We plot net volumes divided by 2 to avoid double-counting trades, which ensures comparability to gross trading volumes (as reported by CRSP).



In a next step, we verify that our conclusions are not driven by large offsetting trades within the aggregated residual investor. Panel a) of Figure 3 plots the ratio of gross volume to net volume for the entire sample (that is, all stocks) and for the sample that includes only shares with institutional ownership above 95%. Importantly, the two lines lie on top of each other for almost the entire sample period, suggesting that differences between gross and net trading volume are not due to unobserved

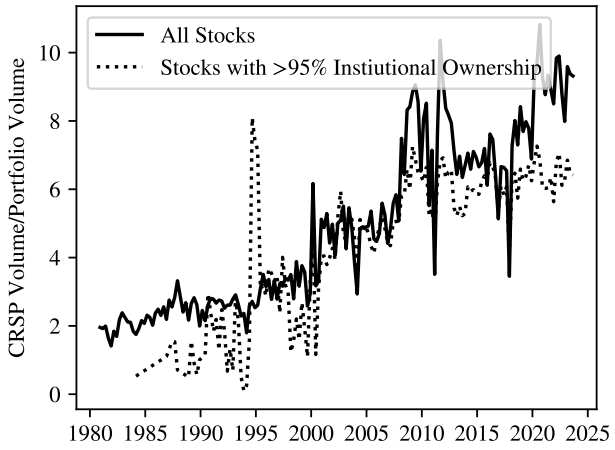
changes in investor portfolios. The low net volumes are also not driven by aggregation across mutual funds within a management company. In Appendix C.1, we disaggregate 13F managers into their constituent mutual funds and ETFs and show that net volumes computed at the fund level are only marginally larger than at the institutional level, suggesting that within-fund-family trading is negligible.

Studies using household data further confirm that households have even smaller net volumes compared to institutional investors. Using portfolio holdings data from households, Gabaix et al. (2025) measure the risk transfer – defined as the percent change in the market risk exposure for a group of investors over a given period, a measure closely related to net volume at the aggregate market level. They find that the quarterly risk transfer is only 0.65% for household groups, significantly smaller than the 6% net volume observed for institutional investors at the aggregate market level (as reported in Figure 9 below).

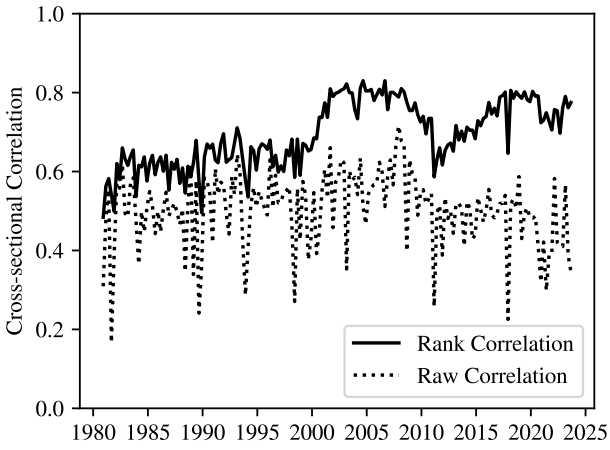
Given the large differences between gross and net volumes, one may wonder whether these measures are fundamentally economically different. Interestingly, Panel b) of Figure 3 suggests they are not. In fact, despite a difference in levels up to a factor of seven, gross and net volumes are remarkably highly correlated in the cross-section. The cross-sectional correlation in ranks is about 80% in recent periods.[13] This high correlation suggests that net and gross volumes are at least to some extent driven by the same fundamental economic primitives. The key difference is that gross trading volume is substantially inflated by round-trip trades that do not contribute to long-term liquidity provision.

Figure 3: **Net Volume versus Gross Volume: Robustness**
Panel a) of the figure plots the ratio of quarterly gross trading volume (CRSP) relative to net volume for all stocks, and stocks with institutional ownership above 95%. Panel b) plots the (rank) correlation of total and net trading volume.

(a) $\frac{\text{Gross Volume}}{\text{Net Volume}}$



(b) Correlation: Total vs. Net Volume



[13]Mechanically, net volume is part of gross volume by construction. However, this cannot explain the high correlation due to the large discrepancy in the level.

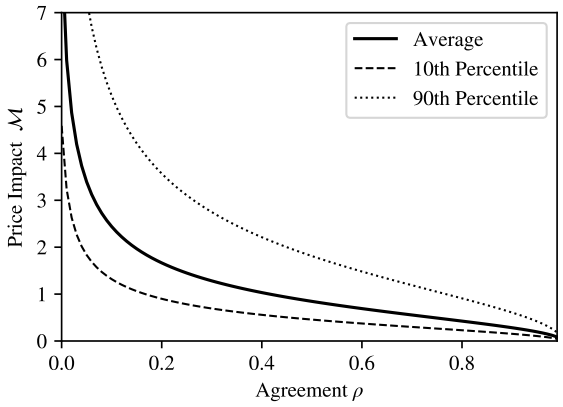
4 The Price Impact Bound

4.1 The Price Impact Bound under Varying Levels of Investor Homogeneity

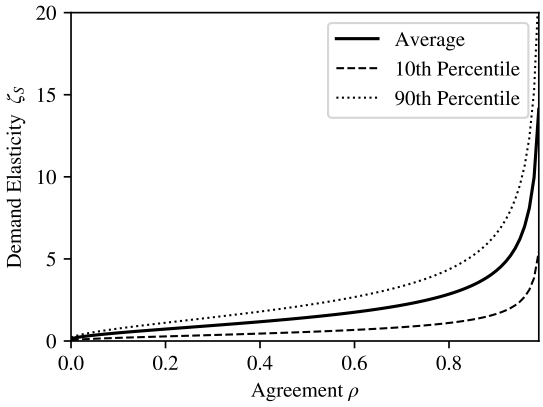
As discussed above, in a first step, we evaluate the price impact bounds without taking a stance on the level of investor homogeneity ρ. In particular, we document the bounds $\mathcal{M}(\rho)$ as a function of ρ for U.S. equities. That is, using stock-level return volatility, σ_p, and net volume, σ_q, constructed as described in the previous section, we compute the bound $\mathcal{M}(\rho)$ for each stock while varying ρ between 0 and 1. Panel a) of Figure 4 plots the distribution of the lower bounds of price impact for individual U.S. stocks. In contrast, Panel b) plots the distribution of the upper bounds on aggregate elasticity, i.e. the inverse of the price impact bounds.

Figure 4: **Price Impact Bounds under Varying Homogeneity ρ**
The figure plots the price impact bound for a given level of investor homogeneity ρ. Panel a) plots the lower bound on the price impact $\mathcal{M}(\rho)$ as a function of ρ, for the average US stock, as well as the top and bottom 10% of stocks with the highest and lowest volatility ratio $\frac{\sigma_p}{\sigma_q}$. The lower bound on price impact bound can be inverted to obtain an upper bound on the aggregate (size-weighted) elasticity. Panel b) plots the upper bound on the aggregate elasticity $\zeta_S(\rho)$ for US stocks.

(a) Price Impact Bound $\mathcal{M}(\rho)$



(b) Elasticity Bound $\zeta_S(\rho)$



As can be seen from the figure, a high volatility-to-volume ratio is consistent with perfectly elastic markets that feature a close-to-zero price impact. However, such a coexistence requires investors be extremely homogeneous, implying that their demand shocks are almost perfectly correlated. Note that the average level of the volatility-to-volume ratio $\frac{\sigma_p}{\sigma_q}$ is about 1. That is, for a 1% demand shock to move prices less than 0.1% ($\mathcal{M} < 0.1$), investor homogeneity must exceed 99%. In other words, the empirical level of the volatility-to-volume ratio can only be reconciled with elastic markets if investors are homogeneous to an extreme, likely unrealistic degree. Put differently, under reasonable levels of investor disagreement, price impact will exceed 0.1%. Generally, the bounds are more stringent as

investor homogeneity decreases ($\rho \to 0$).

4.2 Measuring Investor Homogeneity

Without an explicit measure of investor homogeneity, ρ, the lower bound of price impact cannot be determined. However, as can be seen from Figure 4 the bound is relatively flat when ρ lies between 0.2 and 0.8. In this region, the bound varies predominantly due variation in the volatility-to-volume ratio – at least in our canonical application to the cross-section of U.S. stocks. Panel a) of Appendix Figure E.2 reinforces this conclusion more formally by plotting the partial derivative of the bound with respect to ρ. The derivative is small in absolute terms in a large surrounding neighborhood of $\rho = 0.5$ but grows significantly for extreme degrees of homo-/heterogeneity, i.e., as ρ approaches 1 or 0. The fact that the partial derivative, d, is mostly small implies that a very precise estimate of ρ is required to differentiate between models with $\mathcal{M} = 1$ versus models with $\mathcal{M} = 0.5$. On the other hand, rejecting the null hypothesis that $\mathcal{M} < 0.1$, as implied by most canonical frictionless models, merely requires to show that $\rho < 0.99$. Arguably, this is a much lower hurdle to cross given the extensive literature on heterogeneity in preferences and beliefs among investors. Therefore, rather than trying to provide a precise estimate of ρ, our objective is to establish that ρ is unlikely to be close to either 0 or 1. In this case, the part of Equation (10) which relates to investor heterogeneity, $\mathcal{D} = \sqrt{\frac{1}{\rho} - 1}$, is relatively close to 1 and, thus, the simple volatility-to-volume ratio is as a close proxy for the actual price impact bound.

Unfortunately, common portfolio-based measures of disagreement, such as short interest and active share, cannot directly inform us about investor homogeneity, as these measures are endogenous to prices and thus already contain information about elasticities – the very quantity we seek to speak to. For this reason, we estimate ρ directly from survey data. To that end, we estimate investor homogeneity via stock analyst homogeneity. This approach is almost model-free, as it does not require imposing any specific covariance structure on the underlying demand shocks. However, it does require that the estimated ρ for analysts is "portable" and, thus, reflects well the ρ of investors. Importantly, we do not assume that investors and analysts are the same agents – only that the cross-sectional dispersion in analyst forecasts is a reasonable proxy for the heterogeneity in investors' demand shifts. Notably, because analysts tend to operate within a relatively homogeneous professional environment, and because belief heterogeneity captures only one aspect of broader investor heterogeneity, the limited dispersion in analyst expectations likely understates the degree of heterogeneity among the full set of

investors.

Since analysts submit forecasts across different horizons – from one-quarter ahead to long-term growth rates – and investors care about total discounted cash flows when trading stocks, we estimate analyst homogeneity at different horizons. To be consistent with our theoretical framework, we focus on homogeneity in quarterly updates of forecasts of Institutional Broker Estimates System (I/B/E/S) stock analysts. Specifically, let $\Delta f_{i,t}^h(n)$ denote the update of the earnings per share (EPS) forecast of firm n in period t made by analyst i for horizon h. We then estimate analyst homogeneity $\rho_{EPS}^h(n)$ for each stock n and forecast horizon h as the adjusted R^2 from regressing $\Delta f_{i,t}^h(n)$ on time fixed effects:[14]

$$\Delta f_{i,t}^h(n) = \gamma_t + \epsilon_{i,t}^h(n) \quad \text{for each } n \text{ and } h, \tag{16}$$

where γ_t denotes time fixed effects. We estimate Equation (16) for horizons ranging from one-quarter ahead to three-quarters ahead, as well as long-term growth rates (LTG). The details of the sample construction and estimation procedures can be found in Appendix C.3.

Table 2: Summary Statistics of ρ Estimated from Earnings Forecast Updates
The table reports the distribution of investor homogeneity ρ estimated from analyst forecast updates using Equation (16). For each stock and forecast horizon, ρ is computed as the adjusted R^2 from regressing analyst forecast updates on time fixed effects. 1Q, 2Q, and 3Q refer to one-quarter ahead, two-quarter ahead, and three-quarter ahead earnings per share (EPS) forecasts, respectively. LTG refers to long-term growth forecasts.

Horizon	**# Firms**	**Mean**	**5th Pctl**	**Median**	**95th Pctl**
1Q	754	0.53	0.16	0.56	0.80
2Q	669	0.47	0.11	0.48	0.76
3Q	585	0.41	0.07	0.4	0.75
LTG	366	0.29	0.0	0.26	0.72

Table 2 reports the cross-sectional distribution of $\rho^h(n)$ for different forecast horizons. Intriguingly, analyst homogeneity exhibits a clear term structure across forecast horizons: as the horizon increases, analysts increasingly disagree with each other. This pattern is intuitive – forecast uncertainty grows with the forecasting horizon, and fewer reliable common signals are available for analysts to anchor their expectations. Since a stock's value reflects discounted cash flows across all horizons, investors' demand shocks incorporate innovations to expected cash flows possibly across all horizons. Consequently, estimates derived from forecasts for one quarter and long-term growth can be interpreted as lower and

[14]We first demean forecast updates across time within each analyst, ensuring that the total variation in the regression excludes heterogeneity in average forecast updates. See Appendix C.3 for more details.

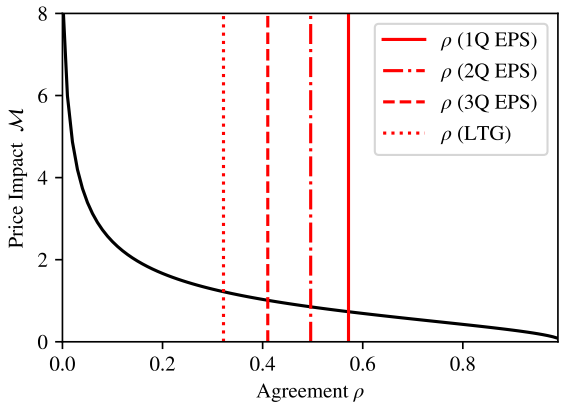
upper bounds of investor homogeneity originating from cash flow expectations.

For stocks in the United States, the average update in one-quarter ahead earnings per share forecasts across analysts explains approximately 53% of the total variation in EPS updates. At this level of investor agreement ($\rho = 53\%$), we obtain an average stock-level price impact of 0.75. In contrast, the average update in long-term growth forecasts explains only 29% of the total variation in LTG updates, implying a price impact of 1.0. Across all measures of investor homogeneity, we rarely observe values of ρ exceeding 80%, suggesting that for the vast majority of stocks price impact exceeds 0.5 and price elasticity falls below 2.

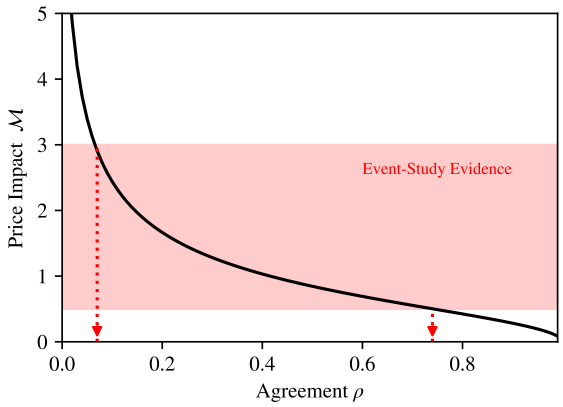
Alternatively, we can compare our estimates of investor homogeneity from I/B/E/S data against the investor homogeneity *implied* by event-study estimates of price impact. To that end, we take the empirical estimates of $\mathcal{M}$ at face value and use Equation (10) to impute ρ. Panel a) of Figure 5 reports average ρ based on survey data along with the stock-level price impact bound as a function of ρ for the average stock. We plot the average ρ obtained from 1, 2, and 3-quarter ahead EPS forecast updates, as well as LTG updates. Panel b) documents the implied ρ based on price impact estimates from the literature. For the range of price impacts found in event studies (such as index inclusions, mutual fund flow-induced trades, and dividend reinvestments) our bound implies that investor agreement ρ should roughly lie between 0.1 and 0.75. Notably, all our estimates from I/B/E/S data are well within this range.

Figure 5: **Investor Heterogeneity: IBES versus Event Studies**
Panel a) plots the average ρ from survey data along with the stock-level price impact bound as a function of ρ for the average stock. We plot the average ρ obtained from 1, 2, and 3-quarter ahead EPS forecast updates, as well as LTG updates. Panel b) plots the investor homogeneity implied from the range of price impacts found in event studies. The dotted lines indicate the implied investor homogeneity by the event-study range.

(a) Average ρ from Survey Data



(b) Implied ρ by Event-Study Literature



4.3 The Price Impact Bound

Next, we apply our estimates of investor homogeneity obtain a lower bound on the price impact for each individual stock as follows:

$$\mathcal{M}_{\text{EPS}}(n) \equiv \frac{\sigma_p(n)}{\sigma_q(n)}\sqrt{\frac{1}{\rho_{\text{EPS}}(n)} - 1} \tag{17}$$

All of the following results are robust to use any of the four forecast-horizon specific estimates of ρ derived from I/B/E/S data in our calculations of the lower bound. However, to maximize the cross-sectional sample size, we rely on ρ estimated based on one-quarter ahead EPS forecasts in our baseline results.

As discussed earlier, the EPS-based price impact bound is imperfect, as it ignores the heterogeneity along many other dimensions. Moreover, for values of ρ in the neighborhood of 0.5, the term $(\sqrt{\frac{1}{\rho} - 1})$ is close to 1 in magnitude and relatively insensitive to the level of ρ. Therefore, we also consider a simplified bound $\tilde{\mathcal{M}}(n)$ defined as the volatility-to-volume ratio, implicitly assuming that $\rho(n) = 0.5$ for all stocks n. Formally,

$$\tilde{\mathcal{M}}(n) \equiv \frac{\sigma_p(n)}{\sigma_q(n)}. \tag{18}$$

Henceforth, we refer to $\tilde{\mathcal{M}}(n)$ as the volatility-to-volume ratio, or simply as the "volatility ratio" when a shorter expression is more convenient. In all our empirical tests, we report results for both $\mathcal{M}_{\text{EPS}}(n)$ and $\tilde{\mathcal{M}}(n)$. Interestingly, $\mathcal{M}_{\text{EPS}}(n)$ contains important incremental information compared to $\tilde{\mathcal{M}}(n)$ when explaining price reactions, despite ρ being measured with noise. Importantly, many other liquidity measures which rely on prices and trading volume such as Amihud (2002) (and the large body of work building on it) are in theory similarly affected by investor homogeneity, ρ. However, this does not directly become evident as many empirical measures of liquidity do not have a micro-founded equilibrium interpretation.

Panel a) of Figure 6 plots the distribution of the price impact bound $\mathcal{M}_{\text{EPS}}(n)$. For the average stock, the lower bound on the price impact is around 1. The top 5% of stocks have bounds exceeding 3.[15] Overall, there is considerable heterogeneity in the bound across stocks which we will explore in the next section. Importantly, the magnitudes of our bounds are consistent with empirical reduced-form evidence from index inclusions (for example, Shleifer (1986)), mutual fund flow-induced trades (for example, Lou (2012)), benchmarking intensity (for example, Pavlova and Sikorskaya (2022)), and

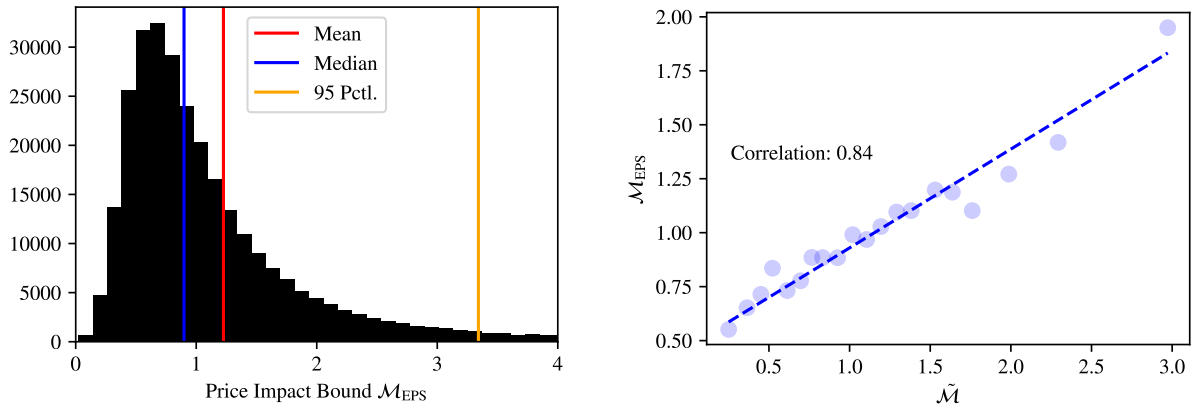
[15]The distribution of the simple volatility ratio, $\tilde{\mathcal{M}}(n)$, is very similar in shape and magnitudes.

dividend reinvestments (for example, Schmickler (2020)). Our bounds highlight that low demand elasticities are not an artifact of unique event studies but are instead a pervasive fact, that can be directly inferred from the high volatility to volume ratio and the amount of investor heterogeneity in the market.

Finally, Panel b) of Figure 6 graphically documents the cross-sectional correlation between $\tilde{\mathcal{M}}(n)$ and $\mathcal{M}_{\text{EPS}}(n)$ is very high at 84%. Relatedly, Appendix Figure E.2 decomposes the cross-sectional variation in $\mathcal{M}_{\text{EPS}}(n)$ and shows that ρ plays a minor role relative to σ_q and σ_p. Put differently, the cross-sectional variation of $\rho_{\text{EPS}}(n)$ is not large enough to dominate the cross-sectional variation in $\frac{\sigma_p(n)}{\sigma_q(n)}$.[16]

Figure 6: **Implied Price Impact for US Equities**
The figure plots the distribution of price impact for the cross-section of US stocks. Panel a) plots the distribution of $\mathcal{M}_{\text{EPS}}(n) \equiv \frac{\sigma_p(n)}{\sigma_q(n)}\sqrt{\frac{1}{\rho_{\text{EPS}}(n)} - 1}$. We use our baseline measure of investor homogeneity ρ extracted from EPS forecast updates ρ_{EPS}. Panel b) plots the correlation between $\mathcal{M}_{\text{EPS}}$ and the simple volatility ratio $\tilde{\mathcal{M}} = \frac{\sigma_p}{\sigma_q}$.

(a) Price Impact Bound $\mathcal{M}_{\text{EPS}}$ (b) Correlation $\mathcal{M}_{\text{EPS}}$ and $\tilde{\mathcal{M}}$



5 Empirical Relevance of the Stock-Level Bounds

Our stock-level bounds are ultimately theoretical constructs that provide lower limits on the expected price impact of investor-specific demand shocks. That is, the bounds are particularly valuable in settings where empirical estimates are unavailable or difficult to obtain. For instance, identifying a source of plausibly exogenous demand shocks to credibly estimate the price impact for broad portfolios, such as the total U.S. equity market, is challenging. Similarly, estimating asset-level price impact is challenging because much of the carefully identified event-study evidence relies on cross-sectional

[16] As discussed in Section 4.2, more formally, the reason for the minor role of investor homogeneity is that the derivative $\frac{\partial \mathcal{M}}{\partial \rho} = -\frac{1}{2}\tilde{\mathcal{M}}\frac{1}{\rho^2}\sqrt{1/\rho - 1}$ is small as long as ρ does not take extreme values, i.e. 0 or 1.

variation and, thus, obtains pooled estimates across assets. However, to trust our model-implied bounds in such a context, it is crucial to verify that the bounds align well with the empirical evidence from settings with credible identification strategies. To that end, we focus on two of the most widely used and verified event studies in empirical asset pricing. Mutual fund flow-induced trades and index inclusions. In particular, we test whether these (plausibly) exogenous demand shifts imply a larger price change for stocks with a higher price impact bound, $\mathcal{M}$.

5.1 Flow-induced trades

Following Coval and Stafford (2007), Lou (2012), and Edmans et al. (2012), flow-induced trades by mutual funds (FIT) have been a widely used source of (plausibly) exogenous variation in demand. We follow the construction of flow-induced trades by Lou (2012) and relegate details to the Appendix C.4. To test whether stocks with higher price impact bounds have higher FIT returns, we interact FIT with $\mathcal{M}_{\text{EPS}}$. We then run panel regressions of quarterly stock returns onto FIT, the interaction of FIT with our bounds, and time fixed effects. Appendix Table E.1 reports corresponding results. As expected, the impact of flow-induced trades is significantly larger for stocks with higher price impact bounds as evidenced by the positive and statistically significant coefficient on the interaction term. Moreover, we sort the stocks into quintiles based on $\mathcal{M}_{\text{EPS}}$ and estimate the flow-induced price impact for each quintile by interacting FIT with quintile dummies. Panel a) of Figure 7 plots the results graphically and Appendix Table E.1 reports the results numerically. In line with our theoretical predictions, price impact estimates increase monotonically moving from the lowest to the highest price impact bound quintile. For example, the flow-driven price impact for the top quintile of stocks is about twice as large in magnitude compared to the bottom quintile.

5.2 Index Inclusions

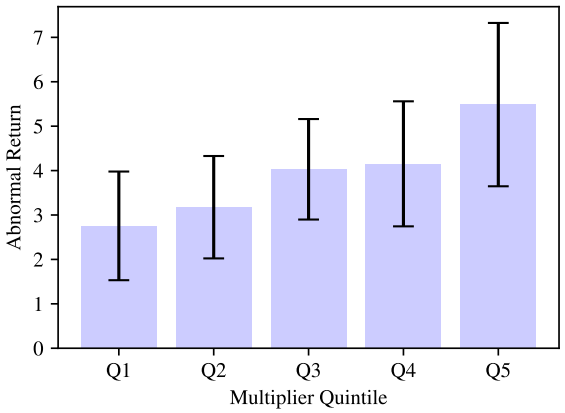
Following Shleifer (1986) and Harris and Gurel (1986), an extensive body of literature investigates the average (abnormal) return around index inclusions and exclusions.[17] Index reconstitutions imply large uninformed demand shifts for the affected securities stemming from passive index trackers who mechanically buy the included and delete the excluded stocks from their portfolios. Relying on the data provided by Greenwood and Sammon (2025) on abnormal event returns and S&P500 reconstitutions, we find an average abnormal event return of 8%. However, there is considerable variation in event returns

[17] Among others, Petajisto (2011), Madhavan (2003), Chang et al. (2015), Pavlova and Sikorskaya (2022)

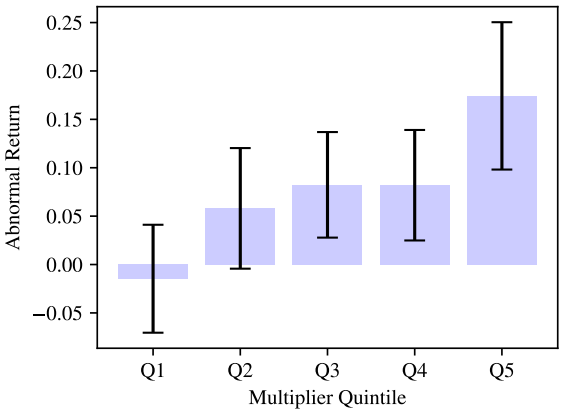
with the cross-sectional standard deviation being equal to 12%. Similar to Section 5.1, we examine whether stocks with higher $\mathcal{M}_{\text{EPS}}$ experience significantly higher abnormal event returns. We find that our price impact bounds are highly statistically significantly related to abnormal event returns. In other words, stocks with high bounds have significantly higher (lower) returns when included (excluded) from the S&P500. As for flow-induced trades, we sort the included and excluded stocks into quintiles by their price impact bound and regress event returns onto the quintiles. Greenwood and Sammon (2025) find that index returns from announcement to effective reconstitution have declined over time, likely because investors increasingly front-run inclusions ahead of the announcement, spreading the effect over a longer window. We therefore focus on the pre-2000 period, when the average index effect was strongest. We also report the results for the whole sample period, which are quantitatively and qualitatively unchanged, but statistically weaker. Panel b) of Figure 7 plots the results graphically and Appendix Table E.2 reports the results numerically. As before, abnormal returns are increasing when moving from the lowest to the highest price impact bound quintile. For example, the abnormal inclusion return for the top quintile of stocks is about 2.5 times as large in magnitude compared to the bottom quintile.

Figure 7: **Validation: S&P500 Inclusions and Flow-Induced Trades**
The figure summarizes the empirical validation of our bounds. Panel a) plots the coefficient of regressing quarterly stock-returns onto flow-induced trades (FIT) interacted with quintile dummies of our price impact bound. Panel b) plots the coefficient of regressing (signed) abnormal event returns during S&P500 index reconstitutions onto quintile dummies of our price impact bound. We report 95% confidence intervals using standard errors clustered by date.

(a) Mutual Fund Flow Induced Trades



(b) S&P 500 Inclusions/Deletions



5.3 Alternative Measures of Price Impact

In this section, we investigate whether the price impact implied by our bounds is different from standard measures of liquidity which rely on gross as opposed to net trading volume. In particular, we use two

alternative measures: the ratio of return volatility to gross-volume $\frac{\sigma_p}{\text{CRSP Vol.}}$ (as opposed to using net volume σ_q in the denominator); and the Amihud (2002) illiquidity measure. Appendix Tables E.3 and E.4 repeat the FIT and S&P500 inclusion regressions for these alternative measures. Importantly, the measures based on gross trading volume do *not explain* the abnormal returns due to demand shocks. In fact, relying on gross rather than net trading volume renders interaction term between FIT and the price impact measure $\mathcal{M}$ insignificant. This suggests that – beyond simply inflating net volume due to round-trip trades – gross trading volume is not suited to measure *long-term* liquidity provision.

In summary, we conclude that our theoretically motivated bounds are empirically relevant. Moreover, the distinction between gross and net volume is particularly important when examining price reactions due to persistent demand shocks.

6 Exploring the Bounds Beyond Event Studies

The previous section documented that our bounds are empirically relevant when measuring long-term price impact of investor-specific demand shocks. Based on this evidence, we next explore the cross-sectional variation of our measures in different settings. To do this, we rely on our simplified measure, $\tilde{\mathcal{M}}$, whenever there is no suitable measure of investor homogeneity ρ available, for example, due to limited time-series variation or lack of estimates of ρ for aggregated portfolios.

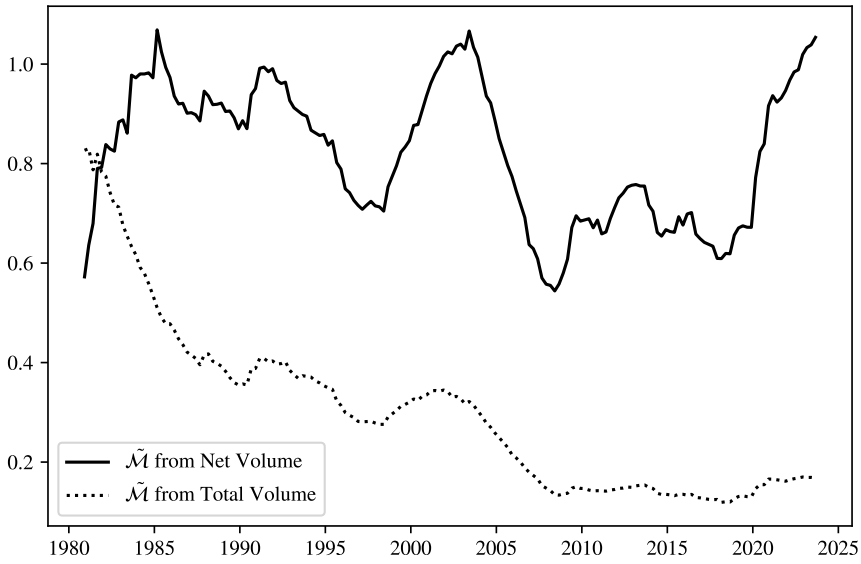
6.1 Secular Trends in Price Impact

First, we examine the evolution of the average stock-level $\tilde{\mathcal{M}}$ during our sample period from 1980 to 2024. To that end, we construct two versions of our simplified bound. First, the solid line in Figure 8 refers the bound as per our theoretical framework (and Equation (18)) and uses net volume in the denominator. Second, the dashed line in the same figure refers to the ratio of return volatility to gross volume, instead.

Quarterly gross volumes have increased monotonically over the past 45 years from 10% in 1980 to 60% in 2024. At the same time, return volatility has remained roughly unchanged over the same time span. Therefore, the price impact bound implied by gross trading volume has continuously declined over our sample period from close to 1 to 0.2. Instead, the level of net volume has remained largely unchanged during our sample period. As a result, the level of the average price impact bound of persistent (quarterly) demand shifts has not materially changed over the same time period. This

stark divergence over time between the two lines in Figure 8 has important implications: Despite soaring trading activity, markets have not become significantly better at absorbing long-term demand shocks. Whereas market participants that have entered since 1980 – such as high-frequency market makers, ETF authorized participants, algorithmic trading firms, and (mobile) retail traders – have likely contributed to the surge in trading volumes, these participants typically engage in short-term strategies do not help absorbing long-term demand shifts. To further substantiate this, Appendix D provides case study evidence on prominent market makers such as Jane Street Capital and Citadel Securities. In fact, Appendix Figure D.1 documents that these two market-making firms jointly account for over 35% of gross trading volume but less than 1% of net volume.

Figure 8: **Secular Trends in Price Impact**
The figure plots the stock-level volatility-to-volume ratio $\tilde{\mathcal{M}}$ for the median stock in the US from 1980 to 2024. We construct $\tilde{\mathcal{M}}$ using both gross trading volume (from CRSP) and net volume to compute σ_q, with σ_p and σ_q based on 5-year backward-looking rolling averages.



6.2 Differences in Price Impact in the Cross-Section of Stocks

In the following we ask the question: Which stocks have higher price impact bounds? To this end, we regress $\mathcal{M}_{\text{EPS}}$ and $\tilde{\mathcal{M}}$ on various stock-specific characteristics such as size (market equity), systematic risk (market beta), momentum (cumulative past returns), book-to-market ratio, dividend to book equity ratio, profitability, and the illiquidity measure from Amihud (2002). Table 3 reports the results.

Table 3: **Heterogeneity in $\mathcal{M}$**
The table summarizes how $\mathcal{M}$ across different stocks. We regress $\mathcal{M}$ and the simplified $\frac{\sigma_p}{\sigma_q}$ on the stock-specific characteristics, log market equity, market beta, momentum, dividend to book equity, profitability, and amihud illiquidity.

	$\mathcal{M}_{\text{EPS}}$			$\tilde{\mathcal{M}}$
	(1)	(2)	(3)	(4)
log(ME)	-0.396***	-0.383***	-0.675***	-0.593***
	(0.017)	(0.017)	(0.033)	(0.029)
β	0.121***	0.159***	0.148***	0.134***
	(0.010)	(0.011)	(0.013)	(0.012)
Cum. Ret.	0.160***	0.151***	0.138***	0.125***
	(0.013)	(0.009)	(0.007)	(0.006)
BM	-0.112***	-0.111***	-0.127***	-0.111***
	(0.013)	(0.012)	(0.012)	(0.011)
$\frac{\text{Dividend}}{\text{BE}}$	0.015	0.023*	-0.028*	-0.026*
	(0.009)	(0.009)	(0.011)	(0.010)
Profit	-0.085***	-0.088***	-0.001	0.001
	(0.011)	(0.010)	(0.011)	(0.010)
Amihud Illiquidity	0.261***	0.265***	0.182***	0.167***
	(0.017)	(0.017)	(0.014)	(0.013)
Date	-	x	x	x
Stock	-	-	x	x
Observations	287895	287895	287895	287895
R^2	0.255	0.283	0.586	0.590
R^2 Within	-	0.250	0.147	0.148

Significance levels: $*$ p < 0.05, $**$ p < 0.01, $***$ p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

First, we find that $\mathcal{M}$ is significantly smaller for larger stocks. That is, a one standard deviation increase in stock size is associated with a 0.13 decline in price impact (t-statistic of 12). This aligns with the view that larger stocks are more liquid, possibly due to more precise and readily available information. Notably, however, this finding contrasts Haddad et al. (2021) and Jiang et al. (2025), who document that large stocks are *less* elastic than small stocks.

Second, stocks with higher market betas exhibit significantly larger price impacts, i.e., a one standard deviation increase in market beta raises the price impact by 0.1. This is consistent with standard CARA-normal intuition: stocks that contribute more to the risk of an arbitrage portfolio are more sensitive to demand shocks (Greenwood (2005), Kozak et al. (2018)).

Third, stocks with stronger past cumulative returns (i.e., momentum stocks) have significantly larger price impacts. This finding aligns with the idea that momentum traders – with upward-sloping demand curves – continue to trade in the direction of the initial price movement, thereby reducing market liquidity and further amplifying price shifts.

Fourth, stocks with higher Amihud (2002) illiquidity have a higher $\mathcal{M}$. Perhaps, this is not sur-

prising as our bounds could be interpreted as low-frequency counterpart to the original Amihud (2002) illiquidity measure. Importantly, however, Amihud (2002) illiquidity does not explain an economically meaningful fraction of our price impact relative to other characteristics. That is, a one standard deviation increase in illiquidity is associated only with an economically relatively small increase in price impact of 0.04. As argued in Section 3.2, gross trading volume (as opposed to net volume) is not well-suited to assess the price impact of long-term demand shifts.

Importantly, all documented patterns are robust – in fact become stronger – when we additionally control for stock fixed effects. Finally, our results also remain unchanged when we use our simplified bounds, $\tilde{\mathcal{M}} = \frac{\sigma_p}{\sigma_q}$, as an independent variable. This further corroborates the fact that our results seem not to be driven by the investor homogeneity parameter which is notoriously difficult to quantify.

6.3 Price Impact for Aggregated Portfolios

Our bounds are particularly helpful for investigating settings for which there is a lack of relevant and exogenous demand shifts, such as the aggregate stock market. Gabaix and Koijen (2021) find that the aggregate stock market is considerably more inelastic than individual stocks. Our simplified bounds are informative about the price impact for aggregated portfolios as they rely only on two simple empirical moments: return volatility and net volume.

Specifically, we compute $\tilde{\mathcal{M}}$ using various portfolio compositions. That is, we start from individual stocks and then successively aggregate to 49 Fama-French industry portfolios, 12 Fama-French industry portfolios, the six portfolios double-sorted on size and book-to-market, three portfolios sorted on size, and, finally, one overall market portfolio. To this end, we first compute return volatility $\sigma_p(g)$ and net volume $\sigma_q(g)$ at these different levels of aggregation. Let $g \subseteq N$ denote the subset of stocks belonging to a given portfolio. Return volatility at aggregation level g is then simply the rolling 5-year standard deviation of the value-weighted portfolio return. For example, for the aggregate stock market, $\sigma_p(g)$ is the standard deviation of the value-weighted return across all stocks. Net volume for aggregation level g is given by

$$\text{NetVol}_t(g) = \frac{\sum_{i=1} \Delta |Q_{i,t}(g)|}{Q_{t-1}(g)}, \tag{19}$$

where $Q_{i,t}(g) = \sum_{n \in g} \Delta Q_{i,t}(n) P_{t-1}(n)$ and $Q_{t-1}(g) = \sum_{i=1}^{I} \sum_{n \in g} Q_{i,t-1}(n) P_{t-1}(n)$. The numerator measures the total dollar flow in and out of portfolio g between $t-1$ and t. The denominator measures the total dollar value of portfolio g as of $t-1$. For example, for the aggregate stock market, the

denominator is given by the total stock market capitalization. As before, we then approximate σ_q as the scaled average of net volume $\sigma_q(g) \approx \frac{\sqrt{\pi/2}}{T}\mathbb{E}[\text{NetVol}_t(g)]$ estimated from 5-year rolling windows.

Figure 9: **Volatility and Volume at different Levels of Aggregation**
The figure plots the volatility of returns $\sigma_p(g)$ and net volume $\sigma_q(g)$ at different levels of aggregation g ranging from individual stocks to the aggregate stock market.

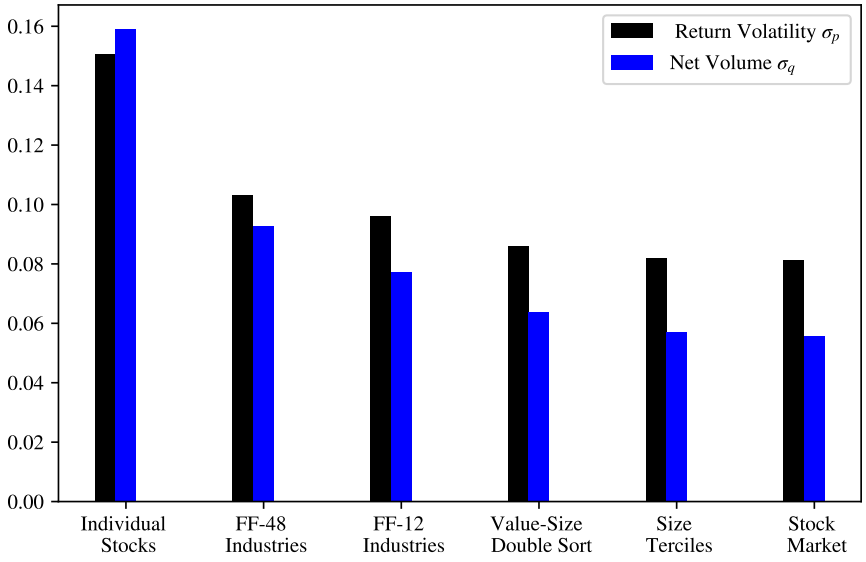
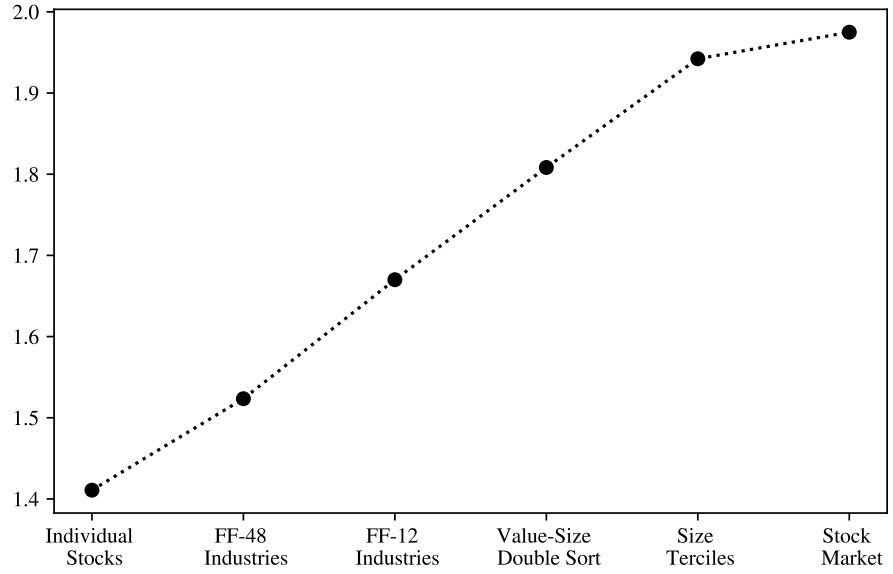


Figure 9 plots our estimates of $\sigma_q(g)$ and $\sigma_p(g)$ at seven different levels of aggregation. At the individual stock level, net volume σ_q is largest. However, as we aggregate stocks into fewer and fewer portfolios, σ_q systematically declines. This pattern is intuitive, as investors' trades in a given stock within a portfolio partly offset each other, which reduces the portfolio net volume. At the same time, return volatility also declines with aggregation. As before, this is intuitive and expected from basic portfolio theory, where diversification reduces idiosyncratic risk. Importantly, however, what matters most for our price impact bounds is the relative speed at which the volatility of returns and net volume decline – ultimately, an empirical question. In the data, return volatility decreases at a lower pace. As a result, $\tilde{\mathcal{M}}$ rises with aggregation as can directly be seen from Figure 10. In fact, the average $\tilde{\mathcal{M}}$ rises from 1.4 to almost 2.0 when moving from individual stocks to the aggregate market.

Figure 10: **Portfolio Level Price Impact**
The figure plots the bound-implied price impact for different levels of aggregation ranging from individual stocks, industries, characteristic portfolios and the aggregate stock market. For each level of aggregation, we plot $\frac{\sigma_p}{\sigma_q}$, which is the price impact implied by $\rho = 0.5$.



6.4 Round-trip Trades and Liquidity Provision

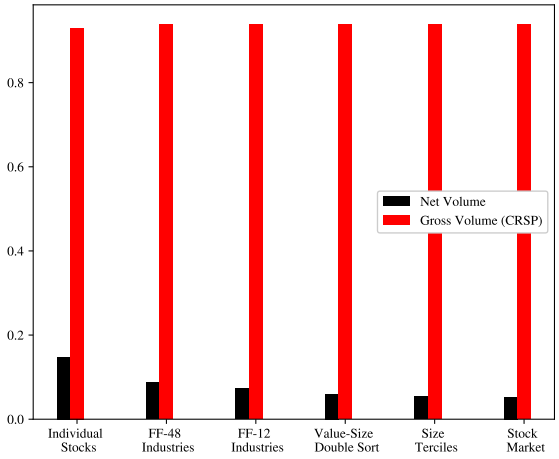
As discussed at length above, while gross volume might be a valuable quantity to examine the market's ability to absorb *transitory* demand shocks, it is not suitable to assess the long-term price impact of persistent demand shifts. In summary, we argue that gross volume overstates long-term liquidity for at least two reasons. First, it takes into account offsetting round-trip trades both *over time* and *across assets*. To further illustrate how relying on gross volume distorts liquidity assessments at lower frequencies, we compare the results from Figures 9 and 10 to the resulting $\tilde{\mathcal{M}}$ when we use gross volume instead of net volume in our calculations. Figure 11 reports the results.

Panel a) illustrates how much larger gross volume is relative to net volume. While net volume declines monotonically with aggregation as discussed above, gross trading volume remains unchanged. This lays bare a key shortcoming of gross volume measures. Without investor-level trade data, it is impossible to identify and control for trades that effectively swap two stocks within the same portfolio and do not provide liquidity to the overall portfolio. Panel b) plots the simplified price impact bound, $\tilde{\mathcal{M}}$, constructed either from gross or net volume. The difference is as striking as intuitive. Gross volume based price impact *increases* with aggregation, whereas net volume based price impact *decreases* with aggregation. That is, liquidity measures based on gross volume can create the misleading impression

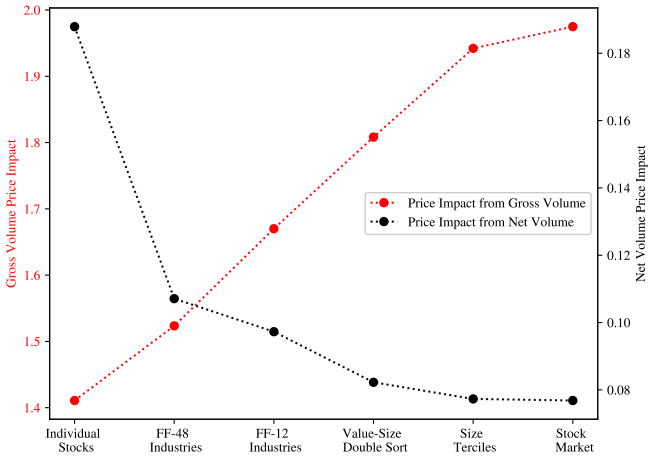
that larger, more aggregated portfolios are more liquid at lower frequencies – that is, trading 1% of their market value incurs a smaller percentage price impact. In contrast, we argue that when using a suitable measure of volume, net volume, the relationship between long-run price impact and aggregation flips.

Figure 11: **Gross vs Net Volume: Implications for Portfolio Price Impact**
The left panel plots gross and net volumes for different levels of aggregation. The right panel plots the simplified price impact bound, $\tilde{\mathcal{M}}$, constructed either from gross or net volume.

(a) Aggregated Volume: Net versus Gross



(b) Implied Price Impact: Net versus Gross



7 Conclusion

This paper reveals a fundamental tension between investor heterogeneity and elasticity: when return volatility is high while trading volume is low, market participants cannot simultaneously be highly heterogeneous in their beliefs and highly elastic in their responses to price changes. We formalize this trade-off through a model-free bound, $\mathcal{M} \geq \frac{\sigma_p}{\sigma_q} \times \sqrt{\frac{1}{\rho} - 1}$, that connects return volatility, net volume, and investor heterogeneity to the price impact of persistent demand shifts. Applied to U.S. equities, our bounds imply substantial price impacts of 0.75 to 1.0 for individual stocks, closely aligning with event study evidence from S&P 500 inclusions and mutual fund flows while traditional gross volume-based liquidity measures fail to explain these price impacts. Despite a five-fold increase in gross trading volume over 30 years, long-term price impact has remained unchanged. Our bounds vary systematically across assets – with larger stocks exhibiting lower price impacts and higher-beta stocks showing greater impacts – and increase substantially with portfolio aggregation, reaching approximately 2.0 for the aggregate stock market. Our bound provides a diagnostic tool for structural models seeking to reconcile volumes and return volatilities, and a sanity check for empirical studies on investor heterogeneity and price impact.

References

Aghaee, A. (2025). The flattening demand curves. *Available at SSRN 4300747*.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets 5*(1), 31–56.

Anshuman, V. R. and S. Viswanathan (2005). Costly collateral and illiquidity. *Duke University working paper*.

Bamber, L. S., O. E. Barron, and T. L. Stober (1999). Differential interpretations and trading volume. *Journal of financial and Quantitative Analysis 34*(3), 369–386.

Banerjee, S. and I. Kremer (2010). Disagreement and learning: Dynamic patterns of trade. *The Journal of Finance 65*(4), 1269–1302.

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics 116*(1), 261–292.

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies 21*(2), 785–818.

Bouchaud, J.-P. (2022). The inelastic market hypothesis: a microstructural interpretation. *Quantitative Finance 22*(10), 1785–1795.

Brennan, M., S.-W. Huh, and A. Subrahmanyam (2013). An analysis of the amihud illiquidity premium. *The Review of Asset Pricing Studies 3*(1), 133–176.

Brennan, M. J., T. Chordia, and A. Subrahmanyam (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics 49*, 345–373.

Brennan, M. J. and A. Subrahmanyam (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics 41*, 441–464.

Bretscher, L., R. Lewis, and S. Shrihari (2025). Investor betas. *Swiss Finance Institute Research Paper*.

Bretscher, L., L. Schmid, I. Sen, and V. Sharma (2025). Institutional corporate bond pricing. *Review of Financial Studies* (forthcoming).

Brunnermeier, M. and L. Pedersen (2009). Market liquidity and funding liquidity. *Review of Financial Studies 22*, 2201–2238.

Campbell, J. Y. and L. M. Viceira (2002). *Strategic Asset Allocation: Portfolio Choice for Long-term Investors*. Oxford University Press.

Chang, Y.-C., H. Hong, and I. Liskovich (2015). Regression discontinuity and the price effects of stock market indexing. *The Review of Financial Studies 28*(1), 212–246.

Chaudhary, M., Z. Fu, and J. Li (2023). Corporate bond multipliers: Substitutes matter. *Available at SSRN*.

Chaudhary, M., F. Zhiyu, and J. Li (2024). Corporate bond multipliers: Substitutes matter. *Available at SSRN*.

Chordia, T., A. Subrahmanyam, and V. R. Anshuman (2001). Liquidity shocks and equilibrium liquidity premia. *Journal of Financial Economics 59*, 32.

Constantinides, G. (1986). Capital market equilibrium with transaction costs. *Journal of Political Economy 94*, 842–862.

Coqueret, G. (2022). Characteristics-driven returns in equilibrium. *arXiv preprint arXiv:2203.07865*.

Couts, S. J., A. S Gonçalves, Y. Liu, and J. Loudis (2024). Institutional investors' subjective risk premia: Time variation and disagreement.

Coval, J. and E. Stafford (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics 86*(2), 479–512.

Dahlquist, M. and M. Ibert (2024). Equity return expectations and portfolios: Evidence from large asset managers. *The Review of Financial Studies 37*(6), 1887–1928.

Darmouni, O., K. Siani, and K. Xiao (2022). Nonbank fragility in credit markets: Evidence from a two-layer asset demand system. *Available at SSRN 4288695*.

Datar, V. T., N. Y. Naik, and R. Radcliffe (1998). Liquidity and asset returns: An alternative test. *Journal of Financial Markets 1*, 203–219.

Davis, S. J. and J. Haltiwanger (1992, August). Gross Job Creation, Gross Job Destruction, and Employment Reallocation*. *The Quarterly Journal of Economics 107*(3), 819–863.

Edmans, A., I. Goldstein, and W. Jiang (2012). The real effects of financial markets: The impact of prices on takeovers. *The Journal of Finance 67*(3), 933–971.

Fang, X., B. Hardy, and K. K. Lewis (2022). Who holds sovereign debt and why it matters. Technical report, National Bureau of Economic Research.

Gabaix, X. and R. S. Koijen (2021). In search of the origins of financial fluctuations: The inelastic markets hypothesis. *Available at SSRN 3686935*.

Gabaix, X., R. S. J. Koijen, F. Mainardi, S. S. Oh, and M. Yogo (2025). Limited Risk Transfer Between Investors: A New Benchmark for Macro-Finance Models.

Garleanu, N. and L. Pedersen (2007). Liquidity and risk management. *American Economic Review 97*, 193–197.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus (2021). Five facts about beliefs and portfolios. *American Economic Review 111*(5), 1481–1522.

Greenwood, R. (2005). Short-and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics 75*(3), 607–649.

Greenwood, R. and M. Sammon (2025). The disappearing index effect. *The Journal of Finance 80*(2), 657–698.

Grossman, S. J. and J. E. Stiglitz (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review 70*(3), 393–408.

Guiso, L., P. Sapienza, and L. Zingales (2008). Trusting the stock market. *the Journal of Finance 63*(6), 2557–2600.

Haddad, V., P. Huebner, and E. Loualiche (2021). How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing. *Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing (April 7, 2021)*.

Han, X., N. L. Roussanov, and H. Ruan (2021). Mutual fund risk shifting and risk anomalies. *Available at SSRN 3931449*.

Hansen, L. P. and R. Jagannathan (1991, April). Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy 99*(2), 225–262.

Harris, L. and E. Gurel (1986). Price and volume effects associated with changes in the s&p 500 list: New evidence for the existence of price pressures. *the Journal of Finance 41*(4), 815–829.

Harris, M. and A. Raviv (1993). Differences of opinion make a horse race. *The Review of Financial Studies 6*(3), 473–506.

Hartzmark, S. M. and D. H. Solomon (2021). Predictable price pressure. *Available at SSRN 3853096*.

Heaton, J. and D. J. Lucas (1996). Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy 104*, 443–487.

Hellwig, M. F. (1980, June). On the aggregation of information in competitive markets. *Journal of Economic Theory 22*(3), 477–498.

Hong, H. and J. C. Stein (2007). Disagreement and the stock market. *Journal of Economic perspectives 21*(2), 109–128.

Honkanen, P., Y. Zhang, and A. Zhou, T (2025). Etf dividend cycles predict money market fund flows and treasury yield changes. *Critical Finance Review Forthcoming*.

Huang, M. (2003). Liquidity shocks and equilibrium liquidity premia. *Journal of Economic Theory 109*, 104–129.

Huebner, P. (2023). The making of momentum: A demand-system perspective. In *Proceedings of the EUROFIDAI-ESSEC Paris December Finance Meeting*.

Investment Company Institute (2025, March). Active and index investing, february 2025. `https://www.ici.org/research/stats/combined_active_index`. Accessed April 18, 2025.

Jansen, K. A. (2025). Long-term investors, demand shifts, and yields. *Review of Financial Studies 38*(1), 114–157.

Jansen, K. A., W. Li, and L. Schmid (2024). Granular treasury demand with arbitrageurs. Technical report, National Bureau of Economic Research.

Jiang, H., L. Zheng, and D. Vayanos (2025). Passive investing and the rise of mega-firms. *Review of Financial Studies forthcoming*.

Jiang, Z., R. J. Richmond, and T. Zhang (2022). Understanding the strength of the dollar. Technical report, National Bureau of Economic Research.

Jiang, Z., R. J. Richmond, and T. Zhang (2024). A portfolio approach to global imbalances. *The Journal of Finance 79*(3), 2025–2076.

Jones, C. (2002). A century of stock market liquidity and trading costs. *Available at SSRN 313681*.

Kandel, E. and N. D. Pearson (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy 103*(4), 831–872.

Kaul, A., V. Mehdrotra, and R. Morck (2000). Demand curves for stocks do slope down: New evidence from an index weights adjustment. *The Journal of Finance 55*(1), 893–912.

Koijen, R. S., F. Koulischer, B. Nguyen, and M. Yogo (2021). Inspecting the mechanism of quantitative easing in the euro area. *Journal of Financial Economics 140*(1), 1–20.

Koijen, R. S., R. J. Richmond, and M. Yogo (2024). Which investors matter for equity valuations and expected returns? *Review of Economic Studies 91*(4), 2387–2424.

Koijen, R. S. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

Koijen, R. S. and M. Yogo (2020). Exchange rates and asset prices in a global demand system. Technical report, National Bureau of Economic Research.

Koijen, R. S. J. and M. Yogo (2025, May). On the Theory and Econometrics of (Demand System) Asset Pricing.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. *The Journal of Finance 73*(3), 1183–1223.

Kvamvold, J. and S. Lindset (2018). Do dividend flows affect stock returns? *Journal of Financial Research 41*(1), 149–174.

Lou, D. (2012). A flow-based explanation for return predictability. *The Review of Financial Studies 25*(12), 3457–3489.

Madhavan, A. (2003). The russell reconstitution effect. *Financial Analysts Journal 59*(4), 51–64.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political economy 111*(3), 642–685.

Pavlova, A. and T. Sikorskaya (2022). Benchmarking intensity.

Petajisto, A. (2011). The index premium and its hidden cost for index funds. *Journal of empirical Finance 18*(2), 271–288.

Schmickler, S. (2020). Identifying the price impact of fire sales using high-frequency surprise mutual fund flows. *Available at SSRN 3488791*.

Schmickler, S. and P. Tremacoldi-Rossi (2022). Spillover effects of payouts on asset prices and real investment. *Available at SSRN 4287300*.

Shleifer, A. (1986). Do demand curves for stocks slope down? *The Journal of Finance 41*(3), 579–590.

Tamoni, A., S. Sokolinski, and Y. Li (2024). Which investors drive anomaly returns and how? *Available at SSRN 4242745*.

Vayanos, D. (1998). Transaction costs and asset prices: A dynamic equilibrium model. *Review of Financial Studies 11*, 1–58.

Wurgler, J. and E. Zhuravskaya (2002). Does arbitrage flatten demand curves for stocks? *The Journal of Business 75*(4), 583–608.

Appendix A Proofs

A.1 Proof to Theorem 1

We prove our main theorem 1 under a weaker assumption than Assumption 1:

Assumption A.1. *Denote $\beta^u_{i,S} \equiv \frac{Cov(u_{i,t},u_{S,t})}{Var(u_{S,t})}$ is the coefficient of regressing the demand shock of investor i, $u_{i,t}$, on the aggregate demand shock, $u_{S,t}$. We have the following regularity condition:*

$$\widehat{Var}^{cs}_S(\frac{\zeta_i}{\zeta_S}) - 2\widehat{Cov}^{cs}_S(\frac{\zeta_i}{\zeta_S}, \beta^u_{i,S}) > 0$$

We first provide the proof to Theorem 1 under the relaxed assumption A.1, and discuss the intuition behind the condition.

Proof to Theorem 1. Given the demand curve equation (1) and price equation (3), we have

$$\Delta q_{i,t} = u_{i,t} - \frac{\zeta_i}{\zeta_S} u_{S,t},$$

Let $\sigma_i^2 \equiv Var(u_{i,t})$ denote the variance of investor i's demand shock, and $\sigma_{iS} \equiv Cov(u_{i,t}, u_{S,t})$ denote the covariance between investor i's demand shock and the aggregate demand shock. The variance of flows and price are:

$$\sigma^2_{q,i} = Var(\Delta q_{i,t}) = \sigma_i^2 - 2\frac{\zeta_i}{\zeta_S}\sigma_{iS} + \frac{\zeta_i^2}{\zeta_S^2} Var(u_{S,t})$$
$$\sigma^2_p = Var(\Delta p_t) = \frac{1}{\zeta_S^2} Var(u_{S,t})$$

The size-weighted average flow volatility is:

$$\begin{aligned}
\sigma_q^2 &= \sum_i S_i \sigma^2_{q,i} \\
&= \hat{\mathbb{E}}^{cs}_S\left[\sigma_i^2\right] - 2\hat{\mathbb{E}}^{cs}_S\left[\frac{\zeta_i}{\zeta_S}\sigma_{iS}\right] + \hat{\mathbb{E}}^{cs}_S\left[\frac{\zeta_i^2}{\zeta_S^2}\right] Var(u_{S,t}) \\
&= \hat{\mathbb{E}}^{cs}_S\left[\sigma_i^2\right] - 2\left(\hat{\mathbb{E}}^{cs}_S\left[\frac{\zeta_i}{\zeta_S}\right]\hat{\mathbb{E}}^{cs}_S\left[\sigma_{iS}\right] + \widehat{Cov}^{cs}_S\left[\frac{\zeta_i}{\zeta_S}, \sigma_{iS}\right]\right) + \left(\hat{\mathbb{E}}^{cs}_S\left[\frac{\zeta_i}{\zeta_S}\right]^2 + \widehat{Var}^{cs}_S\left(\frac{\zeta_i}{\zeta_S}\right)\right) Var(u_{S,t})
\end{aligned}$$

Notice that:

$$\hat{\mathbb{E}}_S^{cs}\left[\frac{\zeta_i}{\zeta_S}\right] = \frac{1}{\zeta_S}\sum_i S_i\zeta_i = 1$$

$$\hat{\mathbb{E}}_S^{cs}\left[\sigma_{iS}\right] = \sum_i S_i Cov\left(u_{i,t}, u_{S,t}\right) = Var(u_{S,t})$$

The expression can be simplified as:

$$\sigma_q^2 = \hat{\mathbb{E}}_S^{cs}\left[\sigma_i^2\right] - Var\left(u_{S,t}\right) - 2\widehat{Cov}_S^{cs}\left[\frac{\zeta_i}{\zeta_S}, \sigma_{iS}\right] + \widehat{Var}_S^{cs}\left(\frac{\zeta_i}{\zeta_S}\right) Var\left(u_{S,t}\right)$$

Under Assumption A.1 that $\widehat{Var}_S^{cs}(\frac{\zeta_i}{\zeta_S}) - 2\widehat{Cov}_S^{cs}(\frac{\zeta_i}{\zeta_S}, \beta_{i,S}^u) > 0$, where we note that $\beta_{i,S}^u = \frac{\sigma_{iS}}{Var(u_{S,t})}$, the condition becomes:

$$\widehat{Var}_S^{cs}\left(\frac{\zeta_i}{\zeta_S}\right) - 2\widehat{Cov}_S^{cs}\left[\frac{\zeta_i}{\zeta_S}, \frac{\sigma_{iS}}{Var(u_{S,t})}\right] > 0$$

which implies $\widehat{Var}_S^{cs}\left(\frac{\zeta_i}{\zeta_S}\right) Var(u_{S,t}) - 2\widehat{Cov}_S^{cs}\left[\frac{\zeta_i}{\zeta_S}, \sigma_{iS}\right] > 0$.

Therefore:

$$\sigma_q^2 \geq \hat{\mathbb{E}}_S^{cs}\left[\sigma_i^2\right] - Var\left(u_{S,t}\right)$$

The ratio of σ_q^2 to σ_p^2 is given as:

$$\frac{\sigma_q^2}{\sigma_p^2} \geq \zeta_S^2\left(\frac{1}{Var(u_{S,t})/\hat{\mathbb{E}}_S^{cs}\left[\sigma_i^2\right]} - 1\right).$$

Using the definition $\rho = \frac{Var(u_{S,t})}{\hat{\mathbb{E}}_S^{cs}[\sigma_i^2]}$ from the main text, we get:

$$\frac{\sigma_q^2}{\sigma_p^2} \geq \zeta_S^2\left(\frac{1}{\rho} - 1\right).$$

Taking the square root and using $\mathcal{M} = \frac{1}{\zeta_S}$, we have the bound:

$$\mathcal{M} \geq \frac{\sigma_p}{\sigma_q} \times \sqrt{\frac{1}{\rho} - 1}$$

□

Remarks. As discussed in the main text, introducing heterogeneity in elasticities can increase the flow volatility for a given level of demand heterogeneity, as different responses to price changes provide

another reason to trade other than heterogeneity in demand.

Assumption A.1 further relaxes the independence assumption in Assumption 1 by allowing for the cross-sectional dependence of data-generating process on the demand shocks and the elasticity, captured by the cross-sectional covariance between elasticity and the correlation with the aggregate demand shock, $\widehat{Cov}_S^{cs}(\frac{\zeta_i}{\zeta_S}, \beta_{i,S}^u)$.

The cross-sectional covariance between the elasticity and the correlation with aggregate shocks also affect the flow volatility. To see this more clearly, note that the condition in Assumption A.1 can also be expressed in terms of the correlation with the change in price:

$$\widehat{Cov}_S^{cs}(\frac{\zeta_i}{\zeta_S}, \beta_{i,S}^u) = \frac{1}{\zeta_S^2}\widehat{Cov}_S^{cs}(\zeta_i, \beta_{i,p}^u)$$

where $\beta_{i,p}^u = \frac{Cov(u_{i,t}, \Delta p_t)}{Var(\Delta p_t)}$ is the regression coefficient of the demand shock $u_{i,t}$ on the change in price Δp_t. Notice that though it is defined as the loading on the price, the causality runs the other way: demand shocks move the price, not vice versa.

All else equal, flow volatility can also be high because investors whose demand shocks move the price more (high $\beta_{i,p}^u$) are also less responsive to price changes ($\widehat{Cov}_S^{cs}(\zeta_i, \beta_{i,p}^u) < 0$), and hence their demand shocks are more manifested in the observed trading. Empirically, large investors, who have larger weights in the aggregation and hence typically are more represented in the aggregate demand shocks, tend to be less responsive to price changes in proportion to their size relative to small investors, often due to trading costs or price impact concerns.

On the contrary, when investors whose demand shocks track the price closer are also more price-elastic, $\widehat{Cov}_S^{cs}(\frac{\zeta_i}{\zeta_S}, \beta_{i,p}^u) > 0$, the opposite channel may dampen the observed flow volatility. Intuitively, their demand shocks are less passed through to the realized trading as they react to the disadvantageous price changes. When this force is overly strong, we may even end up in a pathological equilibrium where investors on average sell when they receive positive demand shocks.[18]

The condition in Assumption A.1 allows for the latter case, but essentially requires that it is dominated by the increase in flow volatility due to the dispersion in elasticities.

[18]To be precise, we may have the empirical moment such that $\hat{\mathbb{E}}_S^{cs}[Cov(\Delta q_{i,t}, u_{i,t})] < 0$. Note that $\hat{\mathbb{E}}_S^{cs}[Cov(\Delta q_{i,t}, u_{i,t})] = \hat{\mathbb{E}}_S^{cs}[\sigma_i^2 - \frac{\zeta_i}{\zeta_S}Cov(u_{i,t}, u_{S,t})]$. Using the equality $\hat{\mathbb{E}}_S^{cs}[\frac{\zeta_i}{\zeta_S}Cov(u_{i,t}, u_{S,t})] = Var(u_S) + \widehat{Cov}_S^{cs}\left[\frac{\zeta_i}{\zeta_S}, \sigma_{i,S}\right]$ as in the proof, we can show that $\hat{\mathbb{E}}_S^{cs}[Cov(\Delta q_{i,t}, u_{i,t})] < 0$ when $\widehat{Cov}_S^{cs}(\frac{\zeta_i}{\zeta_S}, \beta_{i,S}^u) > \sqrt{\frac{1}{\rho} - 1}$.

Appendix B Microfoundations

We start with demand curve representations of portfolio choice under CRRA utility in B.1. We then extend the analysis to a learning-from-price model in B.2.

B.1 CRRA Utility

Consider the portfolio choice problem of an investor with CRRA utility in a two-period model. With log-normal returns, the utility maximization gives the standard portfolio choice formula:

$$\frac{PQ_i}{W_i} = \frac{\mu - R_f}{\gamma_i \sigma_R^2}$$

where W_i is the investor's wealth, γ_i the risk aversion, $\mu \equiv \mathbb{E}\left[\frac{D}{P}\right]$ the expected return, R_f the risk-free rate, and σ_R the return volatility.

We perturb the portfolio-choice problem around a symmetric equilibrium where $\frac{PQ_i}{W_i} = 1$ with first-order log-linearization.[19] We use lowercase letters to denote the log of the uppercase counterpart, and use bar and Δ to indicate the symmetric equilibrium value and the deviation from the original equilibrium, respectively. We have:

$$\Delta q_i \approx -\underbrace{\frac{\bar{\mu}}{\bar{\mu} - \bar{r}}}_{\bar{\delta}} \Delta p + \underbrace{\frac{\bar{\mu}}{\bar{\mu} - \bar{r}} \mathbb{E}\left[\Delta d\right] - \Delta \log \gamma_i - \Delta \log \sigma_R^2}_{u_i} \tag{B.1}$$

In the CRRA model, the demand elasticity $\bar{\delta}$ is determined by the risk free rate and the expected return, which in the equilibrium is further pinned down by the return volatility and risk aversion; the demand shifter u_i comes from different sources, changes in expectations about fundamentals ($\mathbb{E}\left[\Delta d\right]$), changes in risk aversion and uncertainty.

B.2 Learning-From-Price Model à la Hellwig (1980)

We extend the CRRA model to incorporate learning-from-price, adapting the framework from Hellwig (1980).

To focus on the learning-from-price mechanism, we consider a simplified version where demand shocks come only from heterogeneous expectations about dividend changes:

[19] By perturbing around the equilibrium with $\frac{PQ_i}{W_i} = 1$, we simplify the expression by eliminating the change in wealth on the left-hand side.

$$\Delta q_i = -\bar{\delta}\Delta p + \bar{\delta}\mathbb{E}_i\left[\Delta d\right] \tag{B.2}$$

The crucial assumption is that investors form expectations about dividend changes using both a private signal s_i and information extracted from the equilibrium price. We specify the information structure in details later. Here, we postulate that expected dividend changes are formed as a linear combination:

$$\mathbb{E}_i\left[\Delta d\right] = \alpha_s s_i + \alpha_p \Delta p \tag{B.3}$$

where α_s and α_p are equilibrium coefficients that reflect how much weight investors place on their private signals versus price information.

Substituting (B.3) into (B.2) gives us the demand curve with learning-from-price:

$$\begin{aligned}\Delta q_i &= -\bar{\delta}\Delta p + \bar{\delta}\left(\alpha_s s_i + \alpha_p \Delta p\right) \\ &= -\underbrace{\bar{\delta}\left(1-\alpha_p\right)}_{\zeta}\Delta p + \underbrace{\bar{\delta}\alpha_s s_i}_{u_i}\end{aligned} \tag{B.4}$$

The key insight is that learning from prices makes demand less elastic: the effective elasticity $\zeta = \bar{\delta}(1-\alpha_p)$ is smaller than the elasticity $\bar{\delta}$ under rational expectations. When investors observe a price increase, they partly interpret it as conveying positive information about fundamentals, leading them to increase rather than decrease their demand.

To interpret the demand curve in the main text, Equation (B.4) is sufficient. For completeness, below we provide a full characterization of the equilibrium to pin down the coefficients α_s and α_p.

Equilibrium Characterization To fully characterize the equilibrium, we need to determine α_s and α_p. We consider a market with N agents of respective sizes S_i (where $\sum_i S_i = 1$), and eventually take N to infinity. We also consider noise traders who submit orders u_n.

Market clearing requires:

$$\sum_i S_i \Delta q_i = 0$$

From the demand equation (B.4), market clearing implies:

$$\begin{aligned} 0 &= -\zeta \Delta p + \bar{\delta}\alpha_s \sum_i S_i s_i + u_n \\ \Rightarrow \quad \Delta p &= \frac{\bar{\delta}\alpha_s s_S + u_n}{\zeta} \end{aligned} \tag{B.5}$$

where $s_S \equiv \sum_i S_i s_i$ is the size-weighted average signal. This can be rewritten as:

$$\Delta p = \frac{\alpha_s}{1-\alpha_p}\left(s_S + \underbrace{\frac{u_n}{\bar{\delta}\alpha_s}}_{\equiv s_N}\right)$$

where s_N represents the effective "noise signal" from noise trading.

Information Structure and Signal Extraction We assume the fundamental follows:

$$D = \bar{D}\exp\left(\Delta d - \frac{1}{2}\sigma^2_{\Delta d}\right)$$

where $\Delta d \sim \mathcal{N}(0, \sigma^2_{\Delta d})$.

Each investor receives a private signal $s_i \sim \mathcal{N}(0, \sigma_s^2)$ with correlation structure:

$$\operatorname{cov}(s_i, s_j) = \rho\sigma_s^2 \quad \text{for } i \neq j \tag{B.6}$$

$$\operatorname{cov}(s_i, \Delta d) = \beta\sigma_s^2 \tag{B.7}$$

In the limit as $N \to \infty$, the conditional expectation of Δd given signals s_i and the aggregate signal $s_S + s_N$ (a linear function of the price) is:

$$\mathbb{E}\left[\Delta d \mid s_i, s_S + s_N\right] = \frac{\beta\sigma_N^2}{\sigma_N^2 + \sigma_s^2\rho(1-\rho)}\, s_i + \frac{\beta\sigma_s^2(1-\rho)}{\sigma_N^2 + \sigma_s^2\rho(1-\rho)}\,(s_S + s_N) \tag{B.8}$$

where σ_N^2 is the variance of the noise signal s_N. The derivation is provided at the end of this section.

Using the price equation, we can express this conditional expectation in terms of s_i and Δp:

$$\mathbb{E}\left[\Delta d \mid s_i, \Delta p\right] = \alpha_s s_i + \alpha_p \Delta p$$

Matching coefficients, we obtain:

$$\alpha_s = \frac{\beta\sigma_N^2}{\sigma_N^2 + \sigma_s^2\rho(1-\rho)} \tag{B.9}$$

$$\alpha_p = \frac{\sigma_s^2(1-\rho)}{\sigma_N^2 + \sigma_s^2(1-\rho)} \tag{B.10}$$

Substituting back into the demand equation:

$$\Delta q_i = \underbrace{\bar{\delta}\frac{\beta\sigma_N^2}{\sigma_N^2 + \sigma_s^2\rho(1-\rho)}}_{u_i} s_i - \underbrace{\bar{\delta}\frac{\sigma_N^2}{\sigma_N^2 + \sigma_s^2(1-\rho)}}_{\zeta} \Delta p$$

The final elasticity expression reveals the trade-off inherent in learning from prices. On one hand, when private signals are less correlated across investors (low ρ), more new information can be extracted from the price, making the market more inelastic. On the other hand, when noise trader flows are larger (high σ_N^2), the price becomes a less precise signal, making the market more elastic.

Derivation of the conditional expectation formula

Proof. The signal covariance matrix is given as (treating each i as infinitesimally small):

$$\Sigma_s = \text{var}\left(\begin{bmatrix} s_i \\ s_S + s_N \end{bmatrix}\right) = \begin{bmatrix} \sigma_s^2 & \rho\sigma_s^2 \\ \rho\sigma_s^2 & \rho\sigma_s^2 + \sigma_N^2 \end{bmatrix}$$

To compute the (2,2) entry, notice that:

$$\begin{aligned} \text{var}(s_S) = \text{var}\left(\sum_i S_i s_i\right) &= \sigma_s^2\left(\sum_i S_i^2 + \sum_{i\neq j} S_i S_j \rho\right) \\ &= \sigma_s^2\left((1-\rho)\sum_i S_i^2 + \rho\sum_i\sum_j S_i S_j\right) \\ &= \sigma_s^2\left(\rho + (1-\rho)\mathcal{H}\right) \end{aligned}$$

where $\mathcal{H} = \sum_i S_i^2$. Taking the limit as $N \to \infty$, we have $\mathcal{H} \to 0$, so $\text{var}(s_S) = \rho\sigma_s^2$.

The covariance between signals and Δd is:

$$\text{cov}(\Delta d, s_i) = \beta\sigma_s^2$$

$$\text{cov}(\Delta d, s_S + s_N) = \beta\sigma_s^2$$

Thus $\Sigma_{s,\Delta d} = \begin{bmatrix} \beta\sigma_s^2 \\ \\ \beta\sigma_s^2 \end{bmatrix}$.

The conditional expectation is given by $\Sigma_{s,\Delta d}^T \Sigma_s^{-1} \begin{bmatrix} s_i \\ \\ s_S + s_N \end{bmatrix}$. Computing this yields the formula in the main text. □

Appendix C Data Construction Details

C.1 Flow Measures

Quarterly trades $\Delta Q_{i,t}(n)$ and changes in shares outstanding $\Delta\bar{Q}_t(n) = \bar{Q}_t(n) - \bar{Q}_{t-1}(n)$ are adjusted for stock splits in quarter t. We construct trades by the residual investor as $\Delta Q_{0,t}(n) = \Delta\bar{Q}_t(n) - \sum_{i=1}^{I} \Delta Q_{i,t}(n)$. All results in the paper are robust to omitting the residual sector and constructing $\bar{Q}_t(n)$ (and the corresponding size weights) as the sum of institutional shares held. However, we prefer the construction of the residual sector as this effectively accounts for the trades by the institutional sector as a whole, which is otherwise omitted. Furthermore, scaling by institutional shares held leads to some large outliers for smaller stocks that are held by very few institutions. Quarterly trades in percent are denoted by $q_{i,t}(n) = \frac{\Delta Q_{i,t}(n)}{Q_{i,t-1}(n)}$. To reduce the effect of outliers, we also use the Davis-Haltiwanger growth rate 1992, following Gabaix and Koijen (2021) $q_{i,t}(n) = \frac{2(Q_{i,t}(n) - Q_{i,t-1}(n))}{Q_{i,t}(n) + Q_{i,t-1}(n)}$. The results are robust to either definition. When using net volume as the $\mathcal{L}_1$ approximation of flow volatility (the size-weighted variance of $q_{i,t}(n)$), there is no need to express trades in percent, as net volume sums raw trades $\Delta Q_{i,t}(n)$ relative to supply. This makes net volume a more robust estimator, less sensitive to outliers, and the treatment of extensive versus intensive margin trades.

C.2 Net Volume at the Fund Level

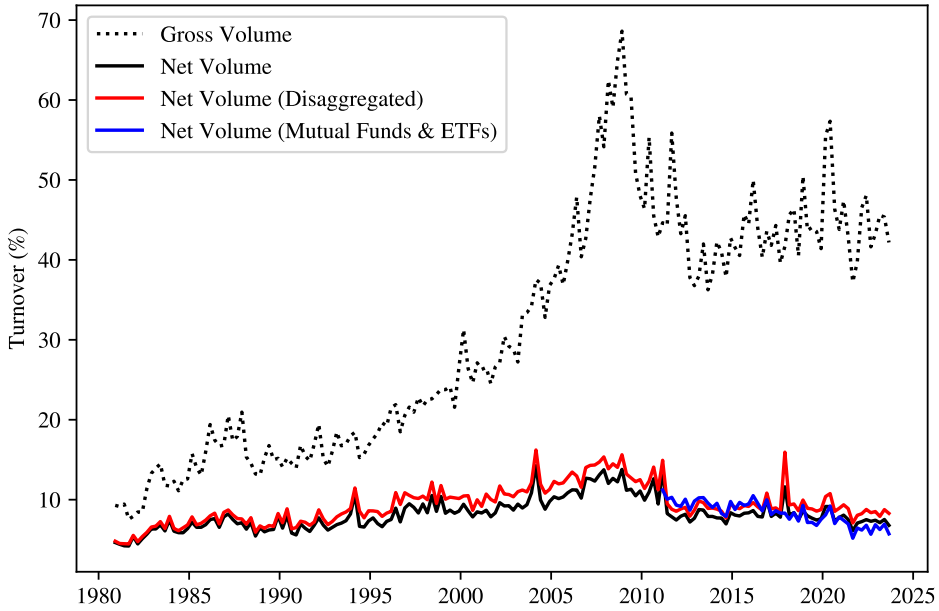
In the main text, we compute all net volumes at the 13F institution level to ensure comprehensive coverage. However, for asset managers with multiple subsidiary funds, institutional-level net volumes exclude intra-family transactions, which may potentially explain why net volumes are smaller than gross volumes. This section uses disaggregated mutual fund holdings data to demonstrate that netting effects from within-institution aggregation are negligible.

We disaggregate fund families in the 13F institutional holdings data (S34 file) using Thomson Reuters mutual fund holdings data (S12 file). Using the S12-S34 link table, we match mutual fund holdings to their corresponding asset managers in the 13F data. For asset managers whose total holdings exceed the sum of their subsidiary fund holdings, we construct a residual entity representing the difference between institutional and mutual fund holdings. We retain institutions in the 13F data that are not matched to any mutual fund. We then compute net volume from this merged dataset using the same methodology as in the main text.

As an additional validation, we construct fund-level net volume using an alternative source: the CRSP Survivor-Bias-Free US Mutual Fund Database, which provides comprehensive coverage of mutual funds and ETFs but excludes other investor types. Since these funds account for a smaller share of market ownership than the broader 13F universe, we normalize net volume within the dataset—dividing net trading activity by the total shares held by all CRSP funds, rather than by shares outstanding. Normalizing by shares outstanding would yield much smaller net volumes and render them non-comparable to those based on 13F data.

Figure C.1 compares net volume measures computed using these two approaches with our baseline institutional-level measures. The red line shows the net volume computed from the disaggregated 13F data using S12 mutual fund holdings files. The blue line presents net volume computed from the CRSP Survivor-Bias-Free US Mutual Fund Database. Despite being computed from different data sources and aggregation levels, the baseline institutional net volume are very close to the fund-level measures, confirming that netting effects from within-institution aggregation are negligible.

Figure C.1: **Net Volume at the Fund Level**
The figure compares net volume measures at the 13F institutional level with net volumes computed at the fund level. *Net Volume* (in black) shows the baseline net volume computed from 13F institutional holdings data. *Net Volume disaggregated* (in red) presents net volume computed from 13F data disaggregated using Thomson Reuters S12 mutual fund holdings files. *Net Volume Mutual Fund & ETFs* (in blue) presents net volume computed from the CRSP Survivor-Bias-Free US Mutual Fund Database, normalized within the dataset.



C.3 Measuring Investor Homogeneity from I/B/E/S

We measure investor homogeneity using analyst forecast data from I/B/E/S, leveraging the idea that the cross-sectional distribution of analyst beliefs serves as a proxy for the cross-sectional distribution of investor demand. This section details the sample construction and methodology for estimating homogeneity parameter ρ.

C.3.1 Data Sources and Sample Selection

We obtain analyst earnings forecasts from the I/B/E/S Detail History database (`ibes.det_epsus`). We only use S&P 500 constituent firms to ensure sufficient number of forecasts. We then link I/B/E/S tickers to CRSP identifiers through a multi-step process: first matching I/B/E/S tickers to Compustat's `gvkey` using the security linking table (`comp.security`), then connecting `gvkey` to CRSP's `permno` through the CCM linking table using link types `LU` and `LC`. Finally, we filter for forecasts made while firms were S&P 500 constituents using historical index membership data.

We focus on two types of forecasts:

- **Quarterly Earnings-per-Share (EPS) forecasts** (FPI codes 6, 7, 8, 9): Representing 1- through 4-quarter ahead EPS forecasts;
- **Long-term growth (LTG) forecasts** (FPI code 0): Representing long-term earnings growth rates.

C.3.2 Construction of Forecast Updates

We identify forecasters at the institution level (`estimator`, brokerage house or sell-side institution), to be consistent with the holdings data which is also at the 13F institution level.

For each forecaster-firm pair, we track how forecasts evolve over time:

EPS Forecast Updates: For quarterly EPS forecasts, we track how forecasters update their forecasts for a specific earnings announcement as it approaches. Each forecast target is uniquely identified by the firm and fiscal period end date (`fpedats`), with the actual earnings released on `anndats_act`. We define the forecast horizon as the number of days between when a forecast is made (`anndats`) and when actual earnings are released (`anndats_act`), converted to quarters by dividing by 90. We retain forecasts made within 400 days of the actual release and round horizons to the nearest quarter with a 30-day tolerance window. When multiple forecasts exist for the same forecaster-target-horizon combination, we select the earliest forecast.

Denote the forecasted EPS by forecaster i at time t for firm n and horizon h as $f^h_{i,t}(n)$. Updates are then calculated as percentage changes between consecutive horizons for the same target:

$$\Delta f^h_{i,t}(n) = f^h_{i,t}(n) - f^{h+1}_{i,t-1}(n)$$

By construction, $f^h_{i,t}(n)$ is around 90 days later than $f^{h+1}_{i,t-1}(n)$, matching the frequency of holdings data.

LTG Forecast Updates: Long-term growth forecasts differ from EPS forecasts as they lack a specific target date and thus no natural horizon measure. For these forecasts, we track quarter-to-quarter changes by assigning each forecast to a quarter based on its announcement date (`anndats`). To avoid partial quarter effects, forecasts made 45 or more days into a quarter are assigned to the following quarter. For each forecaster-firm-quarter combination, we retain only one forecast (the earliest if

multiple exist). Updates are then calculated as simple differences (not percentages) between consecutive quarterly LTG forecasts:

$$\Delta f_{i,t}^{LTG}(n) = f_{i,t}^{LTG}(n) - f_{i,t-1}^{LTG}(n)$$

where $f_{i,t}^{LTG}(n)$ is the long-term growth forecast by forecaster i in quarter t for firm n.

C.3.3 Estimation of Homogeneity Parameter ρ

Following our theoretical framework, we estimate forecaster homogeneity $\rho(n)$ as the adjusted R^2 from regressing individual forecast updates on time fixed effects. Specifically, For each firm n and forecast type (EPS at horizon h or LTG), we then estimate:

$$\Delta \hat{f}_{i,t}^{h}(n) = \overline{\Delta f_{t}^{h}(n)} + \epsilon_{i,t}^{h}(n) \quad \text{for each } h \in \{1, 2, 3, LTG\}$$

where $\Delta \hat{f}_{i,t}^{h}(n) = \Delta f_{i,t}^{h}(n) - \overline{\Delta f_{i}^{h}(n)}$ are the demeaned forecast updates within each forecaster-horizon-firm combination, and $\overline{\Delta f_{t}^{h}(n)}$ are time fixed effects. The adjusted R^2 from this regression captures the proportion of forecast update variation explained by common time effects, serving as our measure of homogeneity $\rho_{EPS}^{h}(n)$.

We use adjusted R^2 as opposed to the original R^2, as the latter can incur a large bias when the number of forecasters is small: When there are only N forecasters, the expected raw R^2 will be around $\frac{1}{N}$ even with completely independent forecasts (hence the population R^2 is 0), while the adjusted R^2 have an expectation of 0 in this case. However, the adjusted R^2 can be negative in the sample. In these rare cases (mostly occur in the LTG forecasts when number of forecasters is small), we truncate the adjusted R^2 at 0.

To further reduce noises due to unbalanced panels, we apply the following filters before estimating $\rho_{EPS}^{h}(n)$: For each firm-horizon pair in quarterly EPS forecasts, we drop forecasters with less than 5 periods of forecast updates, and drop periods with less than 5 forecasters per firm-horizon combination. We repeat this filter iteratively until no more forecasters or periods can be dropped. The LTG forecasts are more sparse, hence we lower the threshold for the number of periods of forecast updates per forecaster-firm-quarter combination and the number of forecasters per firm-quarter combination to 4 and 3, respectively. Table C.1 reports the average characteristics of the final sample.

Table C.1: I/B/E/S Average Number of Forecasters and Updates

The table reports average characteristics of the I/B/E/S forecast sample used to estimate investor homogeneity. "N Periods" refers to the average number of time periods with forecasts per firm-horizon pair. "N Forecasters" is the average number of unique estimators covering each firm-horizon pair. "N Updates per Period" is the average number of forecast updates per firm-period. "N Updates" is the total average number of forecast updates per firm. 1Q-3Q refer to one-quarter through three-quarters ahead EPS forecasts, and LTG refers to long-term growth forecasts.

Horizon	**N Periods**	**N Forecasters**	**N Updates per Periods**	**N Updates**
1Q	44.2	26.0	11.5	510.2
2Q	41.3	25.0	10.9	452.4
3Q	37.6	22.6	10.1	379.8
LTG	16.5	5.2	3.5	57.7

C.4 Flow-Induced Trades by Mutual Funds

TBC

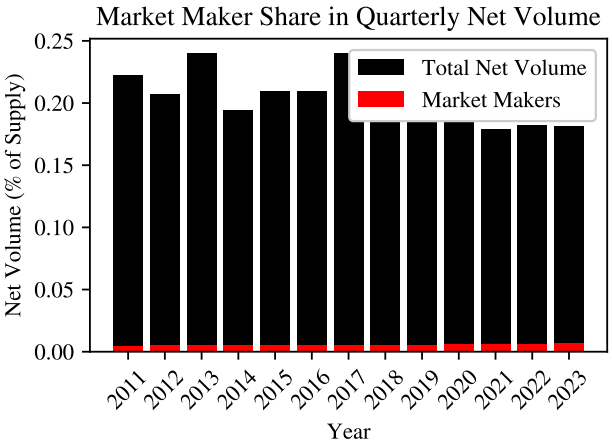
Appendix D Case Study: Citadel and Jane Street

We illustrate the difference between net and total trading volume using a simple example of several market-making firms, which report both quarterly 13F filings (used to compute net volume) and publicly disclosed trading activity (used to approximate total trading volume). Citadel Securities report on their website that they account for 23% of total equity trading volume in the US.[20] Similarly, Jane Street Capital reports that it accounts for 10.4% of equity trading in the US.[21] We compute the total net volume for the largest US market makers that file their quarterly end-of-quarter holdings with the SEC, including Citadel, Jane Street, Virtu, and Two Sigma. Figure D.1 plots their joint share of total net volume from 2011 to 2023. While Jane Street and Citadel alone account for over 30% of total equity trading volume in the US, all of these market makers account for less than 2% of net volume. This underlines that high-frequency market-making firms are less relevant for providing liquidity to persistent demand shifts at longer horizons.

[20] See `https://www.citadelsecurities.com/what-we-do/equities/`, accessed on May, 27, 2025.

[21] See `https://www.bloomberg.com/news/articles/2024-04-17/jane-street-scores-10-6-billion-trading-haul-amid-growth-push`.

Figure D.1: **Case Study: Market Makers and Net Volume**
The figure plots share of total net volume that is driven by Citadel, Jane Street, Virtu, and Two Sigma.



Appendix E Additional Figures and Tables

Figure E.1: **Net Volume versus Flow Volatility**
The figure plots the relationship between flow volatility $\sigma_q(n) = \sqrt{\sum_i S_i(n)\sigma_{q,i}^2(n)}$ and average net volume $\sqrt{\frac{\pi}{2}}\mathbb{E}[\frac{\sum_i |\Delta Q_i|}{Q^*}]$, which are size-weighted averages of $\mathcal{L}_2$ and $\mathcal{L}_1$ norms respectively.

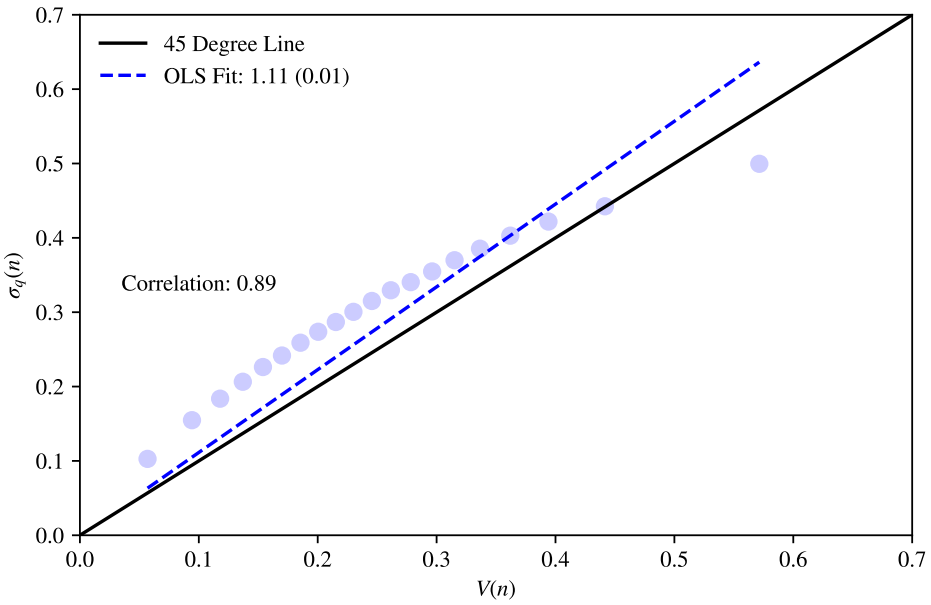
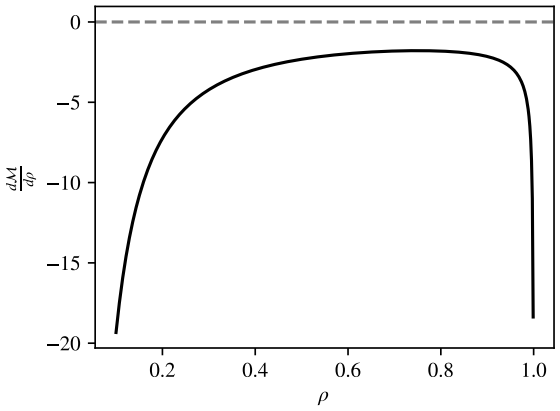


Figure E.2: **Empirical Relevance of ρ**

Panel a) plots the derivative $\frac{\partial \mathcal{M}}{\partial \rho} = -\frac{1}{2}\tilde{\mathcal{M}}\frac{1}{\rho^2}\sqrt{1/\rho - 1}$ as a function of ρ for the average US stock. Panel b) decomposes the variance of $\log \mathcal{M}_{\text{EPS}}$ into its underlying components $\log \sigma_p$, $\log \sigma_q$, $\log \mathcal{D}$ where $\mathcal{D} = \sqrt{1/\rho - 1}$.

(a) $\frac{\partial \mathcal{M}}{\partial \rho}$ for varying levels of ρ



(b) Variance Decomposition of $\log \mathcal{M}_{\text{EPS}}$

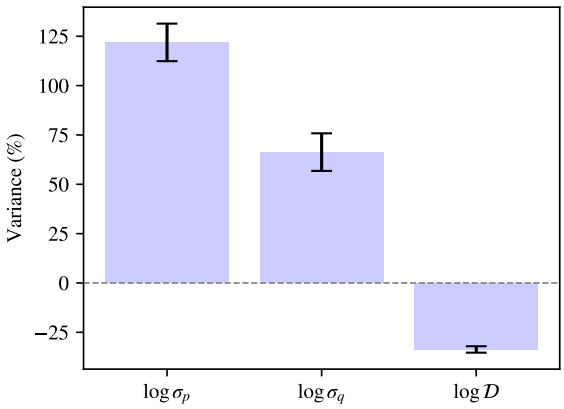


Table E.1: **Validation: Flow-Induced Trades**

The table summarizes the empirical validation of our bounds. We report the Panel coefficient of regressing quarterly stock-returns onto flow-induced trades (FIT) interacted with our bound $\mathcal{M}$, as well as the interaction with quintile dummies of $\mathcal{M}$. Formally, $r_t(n) = \alpha_t + \beta_1 FIT_t(n) + \beta_2 \mathcal{M}_t(n) + \beta_3\big(\mathcal{M}_t(n) \times FIT_t(n)\big) + \epsilon_t(n)$. T-stats are computed using standard errors clustered by date.

	Ret.		
	(1)	(2)	(3)
FIT	3.820***	2.635***	
	(0.510)	(0.636)	
$\mathcal{M}_{\text{EPS}}$		0.004	
		(0.004)	
FIT × $\mathcal{M}_{\text{EPS}}$		1.256*	
		(0.565)	
FIT × $\mathcal{M}_{\text{EPS}}$ quintile: 1			2.755***
			(0.624)
FIT × $\mathcal{M}_{\text{EPS}}$ quintile: 2			3.175***
			(0.588)
FIT × $\mathcal{M}_{\text{EPS}}$ quintile: 3			4.029***
			(0.577)
FIT × $\mathcal{M}_{\text{EPS}}$ quintile: 4			4.152***
			(0.718)
FIT × $\mathcal{M}_{\text{EPS}}$ quintile: 5			5.485***
			(0.938)
Date	x	x	x
$\mathcal{M}_{\text{EPS}}$ quintile	-	-	x
Observations	152862	152862	152862
R^2	0.249	0.250	0.250
R^2 Within	0.004	0.005	0.005

Significance levels: $*$ p $<$ 0.05, $**$ p $<$ 0.01, $***$ p $<$ 0.001. Format of coefficient cell: Coefficient (Std. Error)

Table E.2: **Validation: S&P500 Inclusions**

The table summarizes the price impact of index inclusions and their relationship with our bound. We report the coefficient of regressing (signed) abnormal event returns during S&P500 index reconstitutions onto the bound $\mathcal{M}$. T-stats are computed using standard errors clustered by date. Columns (1)–(3) report results for the full sample period, while columns (4)–(6) restrict the analysis to the pre-2000 subsample.

	Abnormal Return					
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathcal{M}_{\text{EPS}}$		0.058*			0.110*	
		(0.025)			(0.046)	
$\mathcal{M}_{\text{EPS}}$ quintile: 1			0.046***			-0.015
			(0.014)			(0.028)
$\mathcal{M}_{\text{EPS}}$ quintile: 2			0.082***			0.058
			(0.024)			(0.032)
$\mathcal{M}_{\text{EPS}}$ quintile: 3			0.073***			0.082**
			(0.016)			(0.028)
$\mathcal{M}_{\text{EPS}}$ quintile: 4			0.076***			0.082**
			(0.016)			(0.029)
$\mathcal{M}_{\text{EPS}}$ quintile: 5			0.140***			0.174***
			(0.024)			(0.039)
Intercept	0.080***	0.039		0.088***	-0.002	
	(0.008)	(0.021)		(0.016)	(0.039)	
Observations	837	686	686	390	239	239
R^2	0.021	0.036	0.040	0.038	0.091	0.096
Adj. R^2	0.017	0.029	0.027	0.028	0.072	0.061

Significance levels: $*$ p $<$ 0.05, $**$ p $<$ 0.01, $***$ p $<$ 0.001.
Format of coefficient cell: Coefficient (Std. Error)

Table E.3: **Flow-Induced Trading: Alternative Impact Measures**

The table summarizes the coefficient of regressing quarterly stock-returns onto flow-induced trades (FIT) interacted with our bound $\mathcal{M}$, the simplified bound $\tilde{\mathcal{M}}$, the bound constructed from CRSP total volume $\frac{\sigma_p}{\text{CRSP Vol.}}$, as well as Amihud liquidity. T-stats are computed using standard errors clustered by date.

	Ret.			
	$\mathcal{M}_{\text{EPS}}$	$\tilde{\mathcal{M}}$	$\frac{\sigma_p}{\text{CRSP Vol.}}$	Amihud Illiquidity
	(1)	(2)	(3)	(4)
FIT	2.635***	2.505***	3.376***	3.772***
	(0.636)	(0.622)	(0.574)	(0.568)
$\mathcal{M}$	0.004	0.004	0.023	-0.001***
	(0.004)	(0.005)	(0.012)	(0.000)
FIT $\times$ $\mathcal{M}$	1.256*	1.519*	2.352	0.014
	(0.565)	(0.623)	(1.277)	(0.050)
Date	x	x	x	x
Observations	152862	152862	152862	152862
R^2	0.250	0.250	0.250	0.250
R^2 Within	0.005	0.005	0.005	0.005

Significance levels: $*$ p $<$ 0.05, $**$ p $<$ 0.01, $***$ p $<$ 0.001. Format of coefficient cell: Coefficient (Std. Error)

Table E.4: **S&P500 Inclusions: Alternative Impact Measures**

The table summarizes the price impact of index inclusions and their relationship with our bound. We report the coefficient of regressing (signed) abnormal event returns during S&P500 index reconstitutions onto the bound $\mathcal{M}$, the simplified bound $\tilde{\mathcal{M}}$, the bound constructed from CRSP total volume $\frac{\sigma_p}{\text{CRSP Vol.}}$. T-stats are computed using standard errors clustered by date.

	Abnormal Return			
	$\mathcal{M}_{\text{EPS}}$	$\tilde{\mathcal{M}}$	$\frac{\sigma_p}{\text{CRSP Vol.}}$	Amihud Illiquidity
	(1)	(2)	(3)	(4)
$\mathcal{M}$	0.058*	0.114***	0.062	0.008
	(0.025)	(0.027)	(0.046)	(0.004)
log(ME)	-0.000	0.008	0.007	0.003
	(0.013)	(0.011)	(0.013)	(0.014)
β	0.009	0.007	0.018	0.020
	(0.011)	(0.010)	(0.010)	(0.010)
$\frac{\text{Dividend}}{\text{BE}}$	0.011	0.005	0.005	0.008
	(0.008)	(0.007)	(0.007)	(0.008)
Profit	-0.021*	-0.023*	-0.025**	-0.021*
	(0.010)	(0.009)	(0.009)	(0.010)
Observations	686	837	837	666
R^2	0.036	0.049	0.025	0.027
Adj. R^2	0.029	0.043	0.019	0.020

Significance levels: $*$ p $<$ 0.05, $**$ p $<$ 0.01, $***$ p $<$ 0.001. Format of coefficient cell: Coefficient (Std. Error)