Price Multipliers are Larger at More Aggregate Levels

December 31, 2023

Jiacui Li, Zihan Lin

Abstract

We decompose demand imbalances in the U.S. stock market into components at different levels of aggregation and estimate their respective price impacts using a unified approach. The results reveal that the price multipliers form a continuum that is higher at more aggregate levels. Our findings are inconsistent with information-based explanations but are largely consistent with mechanisms based on risk-averse liquidity providers. Our paper proposes a new demand measure for asset-pricing studies, provides support for the "flow-driven" view of aggregate price fluctuations, and bears implications for the modeling of demand-based price effects.

Keywords: Price multiplier, liquidity provision, demand-based asset pricing

JEL classification: G11, G12

Li (jiacui.li@eccles.utah.edu) is at the David Eccles School of Business, the University of Utah. Address: 1655 Campus Center Drive, Salt Lake City, UT 84112. Phone: (401) 688 0584. Lin (zihanl@stanford.edu) is at the Stanford Institute for Computational & Mathematical Engineering. An earlier draft of this paper circulated under the name "Prices Are Less Elastic For Less Diversifiable Demand." The authors are grateful for comments from Yu An, Kerry Back, Bradyn Breon-Drish, David Brown, Scott Cederburg, Bidisha Chakrabarty, Thummim Cho, Carter Davis, Prachi Deuskar, Darrell Duffie, Larry Harris, Da Huang, Timothy Johnson, Mahyar Kargar, Arvind Krishnamurthy, Wenhao Li, Sida Li, Ananth Madhavan (Blackrock), Stefan Nagel, Giang Nyugen, Neil Pearson, Markus Pelger, Andrea Rossi, Mihail Velikov, Xin Wang, Edward Watts, Jun Wu (WRDS), Mao Ye, Jinyuan Zhang, and Yu Zhu as well as seminar and Brown bag participants at Arizona Eller, BYU Marriott, FMA, the Mid-Atlantic Research Conference in Finance, Penn State Smeal, Stanford GSB, the UConn School of Business, UIUC Gies, USC Marshall, Utah Eccles, and UVA Darden. We thank Nimesh Patel and Ivo Welch for sharing S&P index reconstitution data. Jaehee Han, Da Huang, and Yu Zhu provided excellent research assistance.

1 Introduction

Past research has yielded the discovery of many forms of trading flows at various levels of aggregation ranging from the idiosyncratic stock level (e.g. Shleifer, 1986) to more aggregate levels such as style–level fund flows (Teo and Woo, 2004), flows towards sustainable funds (Van der Beck, 2022), etc. How important are these demand changes in shaping asset prices, and how does the answer vary across different levels? This depends on the Gabaix and Koijen (2022) price multiplier M: Increasing demand by \$1 changes an asset's market value by \$M.[1] Thus, a natural question is whether and how the price multiplier varies across levels of aggregation, but there is little consensus in theory: there are models that predict aggregate multipliers to be higher, the same, or lower.[2]

This question regarding price multipliers and aggregation levels has yet to be addressed using a *unifying* empirical approach to estimate and compare price multipliers. Existing studies estimate price effects at only one level at a time and use different data sources and methodologies, making it difficult to compare their results. For instance, index change-based papers only estimate the stock-level price multiplier (e.g. Shleifer, 1986), while the granular instrumental variable approach in Gabaix and Koijen (2022) is used only to estimate the market-level multiplier. In this paper, we simultaneously estimate multipliers at multiple levels with a single demand instrument, thereby rendering the estimates comparable; this also allows us to conduct formal statistical tests for the differences. We find that the multiplier forms a continuum and rises with the levels of aggregation.

We use weekly Lee-Ready signed order flow imbalance as our demand measure. Specifically, for every trade in the U.S. stock market, we first use the Lee and Ready (1991)

[1] A similar view is expressed in Brunnermeier, Farhi, Koijen, Krishnamurthy, Ludvigson, Lustig, Nagel, and Piazzesi (2021): "identifying demand elasticities is a central goal in this literature." In an equilibrium model, a price multiplier is the reciprocal of a demand elasticity.

[2] Theories based on asymmetric information generally predict smaller multipliers at higher levels of aggregation, arguing that asymmetric information is more pertinent at idiosyncratic levels (e.g. Subrahmanyam, 1991). The logistic demand specification in Koijen and Yogo (2019) and papers that follow them assume equal price multipliers across different levels. On the other hand, mechanisms based on investor risk aversion typically predict higher multipliers at higher levels of aggregation (e.g. Petajisto, 2009; Kozak, Nagel, and Santosh, 2018). Please see section 5.1 for additional details.

algorithm to infer its direction as an aggressive buy or an aggressive sell, and then aggregate the trading imbalances during each week and normalize it by lagged shares outstanding. Because this demand measure makes use of all aggressive trading in the stock market, it provides sufficient variation for estimating price multipliers at multiple levels. While prior work has seen this primarily as a "microstructure" trading measure, we show that it has persistent price effects at weekly to quarterly horizons, making it relevant for explaining longer-term variations in asset prices.

To study price impacts at different levels, we decompose the demand measure into components with decreasing levels of aggregation,

$$\text{Demand}_{i,t} = \text{Demand}_{i,t}^{\text{Coarse style}} + \text{Demand}_{i,t}^{\text{Granular style}} + \text{Demand}_{i,t}^{\text{Idiosyncratic}} \tag{1}$$

where the first component corresponds to common demand shocks in coarse style portfolios, the second corresponds to granular style portfolios, and the last corresponds to residual stock-level demand. In our main specification, we use size and book-to-market (BM) characteristic sorts to conduct the style-based decomposition. To infer price multipliers, we regress contemporaneous stock returns on these decomposed demand components and find that price multipliers rise as the aggregation level rises, from 1.73 at the idiosyncratic level to 3.12 at the granular style level and 6.98 at the coarse style level. In other words, buying 1% of shares outstanding moves prices by 1.73%, 3.12%, and 6.98%, respectively, at these levels. The differences between multipliers are statistically significant.

Our finding that multipliers are larger at more aggregate levels is robust. The finding does not depend on using size and BM characteristics, as we obtain similar findings when using demand decomposition based on alternative stock characteristics, as well as when using a data-driven approach that flexibly incorporates many stock characteristics to decompose demand. We also verify that our finding is robust across sub-periods throughout the sample of 1993 to 2022, which helps alleviate concerns about the applicability of the Lee-Ready algorithm to the later sample period. Overall, our finding is statistically robust, but there is uncertainty regarding the mechanism. Do the regression results imply that demand impacts prices? If so, can the results be interpreted as demand effects, or are they contaminated by private information in the demand measure?

We examine the mechanism in three steps. In the first step, we make sure that demand indeed has price effects. For this purpose, we test a concern about reverse causality: perhaps demand does not affect prices but merely chases past returns, and this spuriously shows up as a "contemporaneous" relationship when using lower frequency data (e.g. Schmickler, 2020). Based on this alternative hypothesis, if we were to re-estimate price multipliers using higher frequency data, we would expect the price multipliers to fade away. However, we verify that our results are unaffected when we estimate price impacts at daily, hourly, and 10-minute frequencies. Combined with the prevailing consensus from the microstructure literature according to which order-flow imbalances do impact prices, we believe our result likely reflects movements in prices caused by demand.

Even if demand has price effects, to interpret our estimates as price multipliers we still need to make sure that the price effects are not driven by information about cash flows (e.g. Kyle, 1985). In the second step, we examine this information-based alternative hypothesis. Specifically, we study whether demand predicts future stock earnings above and beyond current market expectations, which we proxy using lagged stock analyst forecasts. The tests reveal that demand has little predictive power over future earnings at all horizons during the next five years. These tests do not lack statistical power: when using cash flow-relevant variables such as stock returns as the independent variable, we find substantial earnings predictability. Demand also does not correlate with contemporaneous releases of earnings or news about a company. Overall, the results are not supportive of the information-based hypothesis.

It is worth clarifying that we do not believe that our demand measure *never* carries cash-flow information. It almost certainly does in some circumstances. However, our findings indicate that the *average quantity* of cash-flow information is insufficient to explain the empirically estimated price multiplier. To further test the information-based explanation, we also collect several indicators that measure the *quantity* of stock- and macro-level news released each week. Under the information-based alternative hypothesis according to which price multipliers arise from information, we should expect our main finding to weaken in subsamples with few news releases. However, across all subsamples with more or less news, our results are virtually unchanged. It is also worth noting that information-based models generally predict smaller price multipliers at more highly aggregate levels (e.g. Subrahmanyam, 1991), which is counterfactual. Overall, these realizations made us conclude that the information-based hypothesis is unlikely to explain our findings.

In the third step, we ask which theories might explain our findings. The structural models in Koijen and Yogo (2019) and follow-up papers predict approximately equal price multipliers at all levels and thus are inconsistent. We argue that models of price impact based on investor risk aversion are the most promising (e.g. Petajisto, 2009; Kozak et al., 2018). In these models, investors are reluctant to trade against price dislocations that are less diversifiable, and this explains why price multipliers rise with the level of aggregation. Having said that, while these models fit the *qualitative* pattern of multipliers, a *quantitative* discrepancy remains. Specifically, these models predict that the stock-level multiplier should be at least an order of magnitude smaller than the style-level one, but we do not find such a large difference in empirical work. We end the discussion with a tentative discussion of possible mechanisms that reconcile this discrepancy.

The main contribution of this paper is using a unified methodology to demonstrate that price multipliers form a continuous spectrum that rises with the levels of aggregation. Combined with evidence reported in previous literature of unsophisticated trading flows at the asset class and style levels, our findings suggest that demand matters more, not less, for shaping prices at more aggregate levels. Regarding empirical quantification of demand-based price effects, our findings suggest that modifications of commonly used logistic demand functions are needed (e.g. Koijen, Richmond, and Yogo, Forthcoming; Chen, Liang, and Shi, 2023). Because these models assume equal price multipliers at all levels, they tend to underestimate the price impact of flows at more highly aggregated levels, such as flows driven by ESG-driven demand tilts (e.g. Van der Beck, 2022). Relatedly, our findings also imply that existing estimates of long–short asset-pricing-factor (anomaly) profit capacities, which are usually estimated using stock-level measures of price impacts (e.g. Korajczyk and Sadka, 2004; Novy-Marx and Velikov, 2016; Ratcliffe, Miranda, and Ang, 2017; Frazzini, Israel, and Moskowitz, 2020), may be overly optimistic.

A secondary contribution involves introducing Lee-Ready signed order-flow imbalance as a demand instrument for asset-pricing studies. This measure has been traditionally thought to operate at the "microstructure" level and has been used primarily at intraday frequencies; the longest horizon examined is daily (Chordia, Roll, and Subrahmanyam, 2002). In this paper, we show that order-flow imbalance does not revert and creates price impacts that last for months. We also show that it contains very little cash-flow information. Combined with the fact that order flow imbalance has ample variation at multiple levels of aggregation, we argue that it is another useful candidate demand measure to be used in conjunction with existing demand instruments such as index inclusion (Shleifer, 1986) and fund flow–induced trading (Lou, 2012).

The rest of the paper proceeds as follows. Section 1.1 summarizes the related literature and discusses how our paper differs from existing studies on demand-based price effects. In Section 2 we introduce our demand measure and examine its properties. In Section 3 we estimate price multipliers and in Section 4 we examine alternative hypotheses. In Section 5 we discuss the relationship between our findings and existing theories. Section 6 concludes.

1.1 Relationship with existing literature

This paper is related to a long literature documenting demand-induced price effects on stocks. Previous studies have studied price effects arising from index additions and deletions

(Harris and Gurel, 1986; Shleifer, 1986; Wurgler and Zhuravskaya, 2002; Chang, Hong, and Liskovich, 2015), mutual fund flows (Teo and Woo, 2004; Coval and Stafford, 2007; Froot and Teo, 2008; Lou, 2012; Huang, Song, and Xiang, 2022; Li, 2022), exchange-traded fund flows (Ben-David, Franzoni, and Moussawi, 2018; Brown, Davies, and Ringgenberg, 2021), payout-induced trading (Schmickler, 2021; Chen, 2022; Hartzmark and Solomon, 2023), futures market order flows (Deuskar and Johnson, 2011), as well as other sources of investor demand variation (Boyer and Zheng, 2009; Parker, Schoar, and Sun, 2020; Ben-David, Franzoni, Moussawi, and Sedunov, 2021a; Peng and Wang, 2021; Liu and Wang, 2021; Li, Pearson, and Zhang, 2021; Gabaix and Koijen, 2022).

Our paper is most related to Gabaix and Koijen (2022) who use the granular instrumental variable approach to estimate the market-level multiplier and show that it is much larger than theoretical predictions. They also argue that market-level multipliers are larger than at the stock level, and propose an explanation based on asset class-level portfolio constraints: investors may be unwilling or unable to flexibly substitute between stocks and bonds. Our paper adds to theirs in three aspects. First, we show that price multipliers form a continuum, rather than a dichotomous "market versus stock" pattern. Second and more importantly, in making the comparison across aggregation levels, Gabaix and Koijen (2022) use a meta-study approach, collecting multiplier estimates from existing papers that use a variety of distinct methodologies. As we discuss further in section 2.2, even at the same level of aggregation, multiplier estimates exhibit sizeable methodology-induced dispersion. In contrast, we use a unified methodology to simultaneously estimate multipliers across different levels which ensures comparability. Finally, in terms of theoretical interpretation, our finding of a continuum of price multipliers is more consistent with an explanation based on investor risk aversion, rather than the asset class-level constraint-based explanation in Gabaix and Koijen (2022) which aims to explain a dichotomous difference between market- and stock-level multipliers.

Our Lee-Ready order-flow imbalance has been used in the microstructure literature to explain price impacts at intraday frequencies. For instance, Breen, Hodrick, and Korajczyk (2002) study the price impact of order-flow imbalance at five-to-thirty-minute intervals, and Hasbrouck and Seppi (2001) document the factor structure in order-flow imbalance at 15-minute intervals. The longest horizon we are aware of appears in Chordia et al. (2002), who demonstrate that daily order-flow imbalance creates price effects at the market level. Our paper differs in two ways ways. First, they exclusively study demand effects at the market level while we compare demand effects across different levels.[3] Second, they study daily price impacts while we focus on longer-term effects; Appendix C.1 shows that the price effects we document persist for months, making our findings relevant for longer-horizon asset pricing purposes. Related to this difference in horizons, they interpret their results as reflecting temporary market-maker inventory imbalances, while we argue that our results are consistent with the idea that investor risk aversion impacts equilibrium prices.

2 Measuring and Decomposing Demand

In this section, we describe our order-flow-based demand measure and our methodology for decomposing it into multiple components at multiple levels of aggregation. We show that demand does not contain information about future cash flows (CFs), which indicates that it captures primarily CF-unrelated trading arising from changes in investor preferences, endowments, etc. We also compare our demand measure with that in previous studies and discuss its benefits and limitations for the purpose of estimating demand effects on prices.

[3] In addition, it is impossible to infer the price multiplier from their specification. This is not a criticism of their paper, as quantitatively estimating price multipliers is not their goal. Specifically, in their main specification for market-level price effects (Table 4), the demand imbalance measure "OIBNUM" is the value-weighted average of the *number* of buy trades minus sell trades. While this is useful for demonstrating the existence of price effects, the coefficient cannot be converted to a price multiplier. The same is true with their other specifications.

2.1 Data and demand measure

Demand data. Our demand measure is Lee-Ready classified order-flow imbalance from the Trade and Quote (TAQ) database for a period running from 1993 through 2022. TAQ is the standard academic source of real-time trading data and contains all quotes and trades executed in the U.S. stock market. We follow the microstructure literature to use the Lee and Ready (1991) algorithm to classify whether a trade is a buy or a sell. To make our findings easy to replicate, our main results are downloaded from WRDS Intraday Indicators, which summarize daily Lee-Ready classified trades at the day-stock level. For the period running from 1993 through 2014, we use seconds-level MTAQ data and Lee-Ready classifications based on interpolated quotes, a specification choice recommended by Holden and Jacobsen (2014), but our results are not sensitive to the use of same-second quotes or previous-second quotes for Lee-Ready classification, both of which are also available from WRDS Intraday Indicators. For the 2015 through 2022 period, we use milliseconds-level DTAQ data. Holden and Jacobsen (2014) discuss the MTAQ and DTAQ datasets in detail.

Our main tests use demand at weekly frequencies. For each stock i in week t, demand is defined as

$$\text{Demand}_{i,t} = \frac{\text{SharesBought}_{i,t} - \text{SharesSold}_{i,t}}{\text{ShareOutstanding}_{i,t-1}} \tag{2}$$

where $\text{SharesBought}_{i,t}$ and $\text{SharesSold}_{i,t}$ are the sum of the total number of shares of stock i bought and sold during week t, respectively, and the denominator is the lagged number of shares outstanding. In auxiliary tests, we also compute demand at other frequencies (e.g. daily) in an analogous manner.

While order-flow imbalances have been used by the microstructure literature (e.g. Breen et al., 2002), our goals led us to construct our measure slightly differently. First, we normalize trades by shares outstanding so that, when we regress returns (dP/P) on the demand measure (dQ/Q), the regression coefficient is calculated in price multiplier units. This departs from most existing microstructure papers that adopt alternative normalizations (e.g. Hasbrouck and Seppi, 2001; Chordia et al., 2002; Chordia and Subrahmanyam, 2004). Second and more importantly, while microstructure studies tend to use order-flow imbalances over shorter windows, such as the five- to thirty-minute intervals in Breen et al. (2002) and the 15-minute intervals in Hasbrouck and Seppi (2001), we examine longer windows because we are interested in effects that are persistent enough for asset-pricing purposes; quickly reverting price impact is irrelevant for explaining longer-term asset-price movements. Unlike what many may assume, order-flow imbalances do not revert quickly (Table A.1) and their price impacts are also persistent (Appendix C.1).

Our demand measure depends heavily on the accuracy of the Lee-Ready algorithm. After comparing Lee-Ready-based classifications against ground-truth data, Chakrabarty, Moulton, and Shkilko (2012) find that, while Lee-Ready classification errors do occur at the trade level, such errors cancel out and misclassification rates converge towards zero when aggregated over a day. Insofar as our main tests are based on daily or slower frequencies, the error rate should be small enough not to substantially impact the inference. We do note, however, that the Lee-Ready algorithm has trouble classifying trades that occur earlier in the trading day when there is no prior price for comparison, which may lead to biases for measuring demand. Thus, we apply a pre-processing step to the demand measure to guard against bias (see details in Appendix A.1), but our subsequent results are robust to directly using demand data without pre-processing.

Other data. We obtain common stock returns and market capitalization data from CRSP. For demand decomposition, we use version 1.1.0 of the stock characteristics data shared by Chen and Zimmermann (2021) on their website. To avoid the concern that results may be driven by microcaps, we remove all stocks with market capitalization below the 20% NYSE cutoff, a common practice in empirical asset-pricing papers (Lewellen et al., 2015, e.g.), but our results are not affected if we do not do so. We use IBES consensus analyst forecasts of earnings per share as proxies for market expectations of future earnings.

2.2 Discussions about the demand measure

We now discuss the rationale for using our demand measure as well as its benefits and drawbacks for estimating price multipliers.

Why is a new demand measure needed? Previous asset-pricing researchers have developed several measures of demand that are plausibly unrelated to CFs, such as trading induced by changes in stock index changes (e.g. Shleifer, 1986). Thus, it is natural to ask why we do not use established demand measures.

While existing demand measures are desirable for many reasons, they typically reflect significant variation only at specific levels of aggregation. For instance, stock-index changes provide significant variation only at the individual stock level (Shleifer, 1986; Harris and Gurel, 1986; Chang et al., 2015, e.g.). Similarly, measures used in Parker et al. (2020), Gabaix and Koijen (2022), Li et al. (2021), and Hartzmark and Solomon (2023) primarily provide market-level variations. However, our goal is to compare price multipliers across levels estimated with a *unified* methodology, which requires a demand measure that exhibits sufficient variation at multiple levels. This is why we turned to our demand measure: it captures enough trading and thus has sizeable variation at all levels.

Why do we insist on using a *unified* empirical methodology to estimate price multipliers across different levels? We do so because the evidence so far indicates that conclusions vary with methodologies. For instance, Table 1 in Gabaix and Koijen (2022) summarizes a subset of price-impact studies using distinct methodologies and shows a wide range of estimates: six studies on stock-level multipliers report estimates ranging from 0.3 to 15, and three papers on market-level multipliers found a range of 1.5 to 6.5. When multiple studies aim at multipliers at the same level, they end up with very different estimates, implying that methodology makes a difference (Menkveld et al., 2023). Therefore, we believe it is important to use a unified methodology to make sure multipliers estimated across levels are comparable.

The advantages of this demand measure. Our demand measure has three main advantages relative to existing ones. First, it is (almost) a *direct* measure of demand which contrasts with many existing measures whereby demand is *estimated*. For instance, to estimate the size of demand movement around the Russell index–change events, Chang et al. (2015) need to estimate the total fund assets indexed to Russell indices and make assumptions about how index funds trade. To cite another example, Hartzmark and Solomon (2023) estimate the amount of investor trading induced by stock dividend payments. In contrast, our measure is computed directly from trades that are reported with accurate quantities.

Second, because TAQ includes all stock trades, our measure captures trading by all market participants. In contrast, most existing measures — in order to focus on well-understood trading behavior — only capture the trading by a subset of investors. For instance, when studying stock index changes, most studies focus on trading behavior on the part of passive funds and do not consider trading by other market participants. However, later work by Pavlova and Sikorskaya (2022) discovered that active funds that are benchmarked to indices also trade in response to index changes, a source of index-induced trading that is missed by previous studies. To the extent that prices are jointly determined by the demand of all market participants, missing trades by other investors can impact the price impact estimates.

Third, our measure can be calculated at high frequency, which is helpful for establishing demand effects, as low-frequency relationships between demand and returns can also arise from return-chasing, as emphasized by Schmickler (2020) when discussing fund flow-induced trading. To take advantage of this feature, while our main tests are run at weekly frequency, in section 4.1 we also examine price impact at daily and intraday frequencies to alleviate concerns about reverse causality.

The limitations of this demand measure. Our measure has two limitations. First, while it captures substantial trading at both the stock and cross-sectional style levels, it misses the bulk of price-relevant trading in S&P futures which prior work has indicated to be the main source of market-level price discovery. Using data in 1993, Chu, Hsieh, and Tse (1999) find that S&P future price movements lead both the S&P exchange-traded fund (ETF) and the underlying stock market. Using a vector-autoregressive framework and data in 2000, Hasbrouck (2003) finds that market-level price discovery predominantly happens in E-mini futures. More recently, using data from 2005 to 2011, Budish, Cramton, and Shim (2015) find that futures lead the S&P ETF in price discovery. Finally, using a unique high-frequency data set, Deuskar and Johnson (2011) find that futures order flow imbalance can explain around 50% of market return variance in a sample from 2006 to 2009.

With help from the authors of Deuskar and Johnson (2011), in Appendix B.4, we find that our market-level demand is positively correlated with S&P futures order flow. Thus, only using our demand to estimate the market-level multiplier will result in a substantial upward bias. Therefore, our demand is only suitable for estimating price multipliers in the cross-section. Therefore, our subsequent price impact regressions always control for time-fixed effects.

Second, by design, our measure only captures trading that is executed aggressively. However, it is known that sophisticated institutions employ optimization algorithms to trade slowly and reduce the use of aggressive orders (Keim and Madhavan, 1995). Therefore, their trades are usually executed as a mixture of aggressive and passive trades, but our measure captures only the aggressive part. As a consequence, we can only estimate the price multipliers associated with aggressive trades which is likely higher than that associated with passive trades. Our main conclusion that "multipliers are larger at more aggregate levels" may also not generalize to passive trades.[4]

[4]Such limitation on generalizability also applies to existing demand effect research as all studies focus on a subset of trades, so the results may not generalize to trades not studied. For instance, index inclusion studies only measure the impact of trades conducted by index funds in response to index changes, so the conclusions may not apply to trades by non-index funds or in other circumstances.

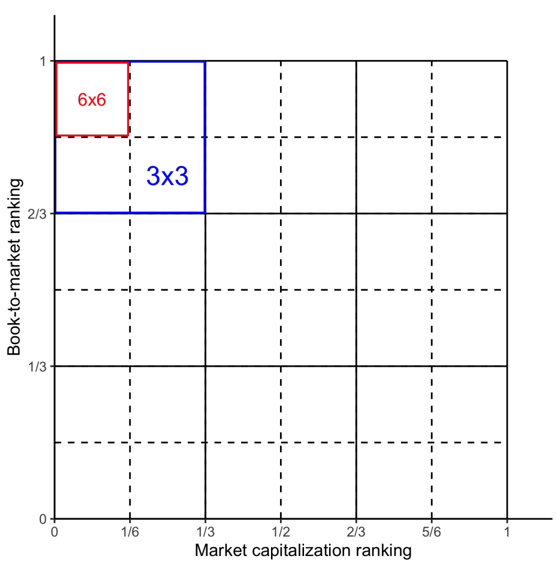
2.3 Demand decomposition using stock characteristics

To study how price multipliers vary across different levels, we start by decomposing demand into multiple components with different levels of aggregation using size and book-to-market (BM) characteristics. We later show that our results are robust to using alternative stock characteristics, but we start with these two because of the prior evidence of investors reallocating capital between size- and BM-based stock styles. Since Barberis and Shleifer (2003) proposed "style investing", researchers have found such behavior in retail investors (e.g. Kumar and Lee, 2006), mutual fund flows (e.g. Teo and Woo, 2004; Li, 2022), and 13F institutions (e.g. Abarbanell, Bushee, and Smith Raedy, 2003; Froot and Teo, 2008).

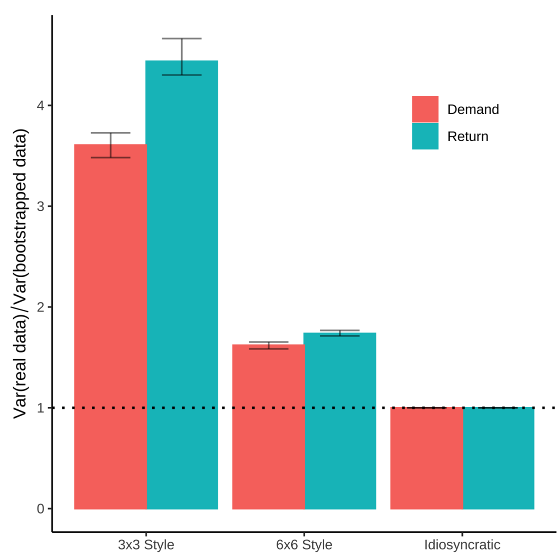
Figure 1. Decomposing demand into components using stock characteristics

We decompose stock-level demand into three components with decreasing levels of aggregation using size and book-to-market (BM) double-sorted portfolios. Plot (a) illustrates the decomposition process. Stocks are double-sorted into nested 3×3 and 6×6 size–BM portfolios. The most aggregate component is defined as the value-weighted average demand in the 3×3 style portfolios (blue); after subtracting the 3×3 style demand component, the value-weighted average demand in the 6×6 portfolios (red) form the next component; further subtracting the 6×6 style demand gives the residual idiosyncratic demand component. For Panel (b) we examine the degree to which demand and return exhibit correlations in size/BM-sorted portfolios. To do so, we compute the ratio of the variance of the actual decomposed demand components to the variance under the bootstrapped null in which demand and returns do not exhibit cross-sectional correlations. A resulting ratio that is above one implies that demand and returns exhibit correlation. The error bars represent 95% bootstrapped confidence intervals.

(a) Size and BM-based demand decomposition



(b) Variance of decomposed components



We decompose demand using Fama-French style portfolio sorts. Specifically, we sort stocks into nested 3×3 and 6×6 size and BM portfolios using NYSE breakpoints in each period, as illustrated in Panel (a) of Figure 1. We then decompose weekly demand for stock i and week t into three additive components:

$$\text{Demand}_{i,t} = \text{Demand}^{3\times 3}_{l1(i),t} + \text{Demand}^{6\times 6}_{l2(i),t} + \text{Demand}^{\text{Idio}}_{i,t} \tag{3}$$

where $l1(i)$ and $l2(i)$ index the 3×3 and 6×6 portfolios, respectively, to which stock i belongs. To compute each demand component, we recursively calculate:

$$\text{Demand}^{3\times 3}_{l1,t} = \sum_{j \in 3\times 3 \text{ style } l1} w^{3\times 3}_{j,t-1} \cdot \text{Demand}_{j,t}, \tag{4}$$

$$\text{Demand}^{6\times 6}_{l2,t} = \sum_{i \in 6\times 6 \text{ style } l2} w^{6\times 6}_{j,t-1} \cdot (\text{Demand}_{j,t} - \text{Demand}^{3\times 3}_{l1(j),t}), \text{ and} \tag{5}$$

$$\text{Demand}^{\text{Idio}}_{i,t} = \text{Demand}_{i,t} - \text{Demand}^{3\times 3}_{l1(i),t} - \text{Demand}^{6\times 6}_{l2(i),t} \tag{6}$$

where $w^{3\times 3}_{j,t-1}$ and $w^{6\times 6}_{j,t-1}$ are lagged market capitalization weights of stock j in its corresponding 3×3 and 6×6 portfolios, respectively. In equation (6), $\text{Demand}^{\text{Idio}}_{i,t}$ is defined as a residual after subtracting the first two components. Therefore, the three decomposed demand components rank from more to less aggregate levels.

These decomposed demand components are easy to interpret. For instance, suppose an investor buys one dollar of Apple stock which, on the day of purchase, happens to be a large cap-growth stock. We decompose the investor's demand into three components: she buys one dollar of the 3×3 large-growth portfolio, one dollar of the portfolio that is long the more granular 6×6 portfolio and short the less granular 3×3 large-growth portfolio, and one dollar of the idiosyncratic component of Apple stock. When we later regress returns on these decomposed demand components, the resulting price multiplier associated with 3×3 style-level demand can be compared to prior estimates of style-level multipliers (e.g. Peng and Wang, 2021; Ben-David, Li, Rossi, and Song, 2022; Li, 2022), and the resulting idiosyncratic-level multiplier can be compared to stock-level studies (e.g. Shleifer, 1986). Koijen and Yogo (2019) assume that these three demand components have identical price multipliers and therefore estimate them together. In this paper, we allow for the possibility that the price multipliers are different.

In Table 1 we report summary statistics for decomposed weekly demand. The third column, in which we report standard deviations, gives a sense of the magnitudes of demand variation at different levels. At the idiosyncratic level, a one-standard-deviation demand shock amounts to buying or selling 0.64% of shares outstanding. Demand is less volatile in aggregate style portfolios but still sizeable: a one-standard-deviation demand shock at the 3×3 and 6×6 style levels amount to buying or selling 0.08% and 0.07% of all shares outstanding in those portfolios, respectively.

Table 1. Summary statistics

In this table we report summary statistics for the weekly sample from 1993 to 2022. The sample excludes all stocks with market capitalization below the NYSE 20% breakpoint. Demand is measured by Lee-Ready classified order-flow imbalance and normalized by the number of shares outstanding. Then, we decompose the demand measure into a 3×3 size-book/market (BM) style component, a 6×6 size-BM style component, and an idiosyncratic residual component, as described in section 2.3. Weekly returns are decomposed in the same way for comparison purposes. Both demand and returns are reported in percentages; in the last two rows, we report the size and BM characteristics. In column (1) we report the average number of stocks in each time period. In columns (2) and (3) we report the means and standard deviations, and in the last five columns, we report the percentile distributions of each variable.

		Obs.	Mean	StDev	Percentiles 5%	25%	50%	75%	95%
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Demand (%)	3×3 style	2,507	0.00	0.08	-0.12	-0.04	0.00	0.03	0.11
	6×6 style	2,507	0.00	0.07	-0.11	-0.04	0.00	0.04	0.11
	Idiosyncratic	2,507	0.00	0.64	-0.85	-0.21	0.00	0.21	0.88
Return (%)	3×3 style	2,507	0.02	1.50	-2.17	-0.65	0.02	0.68	2.30
	6×6 style	2,507	0.00	0.85	-1.24	-0.42	0.00	0.42	1.26
	Idiosyncratic	2,507	0.00	5.92	-8.51	-2.64	-0.15	2.41	8.90
Characteristics	Market cap ($m)	2,507	6,439	31,862	167	405	1,021	3,322	24,523
	Book-to-market	2,507	0.516	0.986	0.060	0.213	0.391	0.636	1.250

Demand exhibits cross-sectional correlations in the size-BM portfolios. To see this, we use a bootstrap procedure to compare the variance of decomposed demand components with that simulated under the null hypothesis that stock-level demand is idiosyncratic. Specifically, in each cross-section, we randomly permute stock characteristics to break cross-sectional correlations and then conduct demand decomposition using bootstrapped characteristics. The red bars in Panel (b) of Figure 1 plot the ratio between the variance of actual demand components and the variance based on bootstrapped data. For instance, the first bar shows that the 3×3 style-level demand variance is 3.61 times as large as the null hypothesis with no correlations. The 6×6 style component exhibits weaker correlation but is still more highly correlated than the bootstrapped idiosyncratic case, and the difference is statistically significant, as shown by the 95% bootstrapped confidence interval in the figure. If demand exhibits no size- or BM-based correlations, we would expect to see a ratio of one, as marked by the horizontal dashed line.

How strong is the demand correlation in the style levels? For an intuitive comparison, we compare it with the degree of correlation in stock returns as it is well-known that returns exhibit strong size- and BM-based correlations (Fama and French, 1996; Kozak et al., 2018). The blue bars in Panel (b) of Figure 1 plot the results of the analogous exercise for stock returns. The comparison reveals that the degree of demand correlation is only slightly lower than that in returns. Overall, these findings reveal that demand has systematic components, consistent with the earlier finding in Hasbrouck and Seppi (2001).

2.4 Does the demand measure capture cash-flow information?

Because our demand measure covers trades by many market participants over a thirty-year sample, it likely captures trades with multiple motivations. It is natural to worry that our demand measure is informative about future stock cash flows (CFs), and that such CF information explains the price multipliers we later estimate. When using future firm-level earnings to proxy for CF, however, we find that our demand measure contains little CF information.

This suggests that the trades captured by our demand measure are driven primarily by CF-unrelated reasons. For institutional investors, CF-unrelated trading can come from the need on the part of fund managers to satisfy purchase or redemption requests in fund flows (e.g. Lou, 2012; Vayanos and Woolley, 2013), to reinvest cash proceeds (e.g. Chen, 2022), etc. For retail investors, CF-unrelated trading can arise from liquidity needs, interest in attention-grabbing stocks (Barber and Odean, 2007), and so on. In Appendix B.1, we report that our demand measure can capture the well-known CF-unrelated trading triggered by S&P stock index changes. When a stock is added to or deleted from the index, our demand measure spikes up or down, respectively.

Does demand predict future earnings? While dividends are the ultimate CF payoff to investors, researchers often use firm-level earnings as a faster-moving proxy (e.g. Vuolteenaho, 2002). Therefore, we test whether our demand measure predicts future earnings per share (EPS). For this exercise, we obtain realized EPS and analyst forecasts of EPS from IBES. We convert our demand data into monthly frequency to match the forecast-updating schedule in IBES. Appendix B.2 provides further details regarding data construction.

We estimate panel regressions to examine whether demand is informative about future EPS beyond what is already reflected in lagged analyst forecasts. For each horizon of $h = 1, ..., 5$ years into the future, we estimate forecasting panel regressions

$$\frac{X_i^h - F_{t-1}(X_i^h)}{P_{i,t-1}} = \sum_{\text{level } l} b_h^l \cdot \text{Demand}_{i,t}^l + \text{Controls}_{i,t} + \epsilon_{i,t} \tag{7}$$

where X_i^h is the h-year ahead annual EPS and $F_{t-1}(X_i^h)$ is the lagged consensus IBES analyst forecast that we use to proxy for pre-existing market expectations. We normalize their difference by $P_{i,t-1}$, the one-month lagged stock price. In the regressions we control for three lags of monthly stock returns as well as monthly and stock fixed effects, and we cluster standard errors by month and stock. Because IBES analysts provide forecasts for earnings up to five years ahead, we examine the predictive power of demand for $h = 1, 2, ..., 5$ years. The main independent variables are the three components of our demand measure.

The regression results are shown in the first five columns of Table 2. Because our demand measure captures aggressive trading, one may suspect it contains near-term CF information. The results reported in columns (1) and (2) show, however, that all three demand components carry little predictive power over the subsequent two annual EPS releases. In columns (3) to (5), which present regressions at 3- through 5-year horizons, we also do not observe evidence of demand predicting future earnings.

To examine nearer-term CF, we also estimate the predictive power of demand over the subsequent four quarterly EPS releases and report the regression results in columns (6) through (9). The results are generally consistent with the results for one-year-ahead annual EPS that indicate no predictive power. In Appendix B.2.1 we show that this finding that demand cannot predict future earnings is robust across sub-samples. Overall, our findings indicate that demand contains little information about *future* CF in the subsequent five years. Later, in section 4.2, we also find that demand is not correlated with *contemporaneous* releases of CF news.

Several questions ensue. First, analyst forecasts may be imperfect proxies for market expectations. Prior studies find that analyst forecasts are often slow to respond to news (e.g. Easterwood and Nutt, 1999; Jegadeesh and Livnat, 2006). However, this would bias in the direction of finding stronger, not weaker, coefficients for demand. Even if demand just reflects stale news, because analyst updates are delayed, we would still expect to find a positive coefficient when demand is used to forecast EPS. Therefore, slow analyst responses cannot explain the *lack* of a relationship.

Second, one may worry that our tests lack statistical power. Specifically, perhaps demand does predict earnings but our regressions are misspecified and thus fail to pick up the relationship. This does not appear likely. The fact that lagged return control variables used in Table 2 all carry significant coefficients indicates that our test has power. Further, when we replace demand with plausibly CF-related variables such as realized earnings surprises or firm-level news sentiment, we also find significant predictive power (Appendix B.2.2).

Table 2. Using demand to predict future earnings

We use panel regressions to examine whether demand predicts future earnings beyond market expectations as captured by previous analyst forecasts. The dependent variables are the differences between realized future earnings per share (EPS) and lagged IBES analyst consensus forecasts divided by the lagged share price. The independent variables are the three components of decomposed demand. In all regressions, we control for three lags of monthly stock returns as well as month and stock fixed effects, and we cluster standard errors by month and stock. For columns (1) through (5) we examine predictive power over annual EPS that will be released 1 to 5 years in the future and for columns (6) through (9) we examine predictive power over quarterly EPS that are 1 to 4 quarters in the future. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

	Future annual EPS					Future quarterly EPS			
	1y ahead	2y	3y	4y	5y	1q ahead	2q	3q	4q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\text{Demand}^{3\times3\text{ style}}_{i,t}$	−0.193**	−0.206	−0.052	0.236	−0.014	−0.092**	−0.177***	−0.128**	−0.119
	(0.088)	(0.164)	(0.156)	(0.257)	(0.269)	(0.042)	(0.050)	(0.062)	(0.073)
$\text{Demand}^{6\times6\text{ style}}_{i,t}$	−0.050	0.000	0.053	−0.073	0.034	−0.026	−0.022	−0.052*	−0.017
	(0.045)	(0.061)	(0.091)	(0.125)	(0.200)	(0.033)	(0.032)	(0.029)	(0.030)
$\text{Demand}^{\text{Idio}}_{i,t}$	0.003	0.007	−0.004	−0.017	−0.009	−0.002	0.000	0.000	−0.002
	(0.005)	(0.005)	(0.008)	(0.013)	(0.014)	(0.002)	(0.002)	(0.003)	(0.004)
$\text{Ret}_{i,t-1}$	0.022***	0.041***	0.041***	0.042***	0.051***	0.008***	0.013***	0.014***	0.013***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
$\text{Ret}_{i,t-2}$	0.016***	0.034***	0.033***	0.033***	0.037***	0.006***	0.011***	0.010***	0.011***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
$\text{Ret}_{i,t-3}$	0.013***	0.027***	0.027***	0.025***	0.027***	0.005***	0.008***	0.008***	0.009***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Stock FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	789,172	699,270	398,390	165,263	102,674	796,500	754,787	711,136	657,816
Within R^2	0.011	0.018	0.012	0.010	0.011	0.007	0.013	0.011	0.010

Sizing price multipliers justified by information. The findings reported in Table 2 indicate that, as far as the next five years of earnings are concerned, CF information cannot justify sizeable price multipliers.

This point can be seen from a simple back-of-the-envelope exercise presented in Appendix B.3, which we summarize here. In a present-value framework, for each level l, the price multiplier justified by earnings information in demand is given by

$$\sum_{h=1}^{\infty} \frac{\tau \cdot b^l_h}{(1+r)^h}, \tag{8}$$

where $\tau \leq 1$ is the ratio of EPS paid out as dividends, $r \geq 0$ is the discount rate, and $\{b^l_h\}_{h=1}^{\infty}$ are the predictive regression coefficients estimated in (7). We now restrict attention to the first five years $h = 1, ..., 5$ for which we have estimates. Even if we assume a payout ratio of $\tau = 1$ and a discount rate of $r = 0$, which will bias the quantification upwards, the point estimates of (8) are -0.229, -0.036, and -0.019 for the 3×3 style, the 6×6 style, and idiosyncratic levels, respectively. This means that the information-justified price multiplier is essentially zero at all levels of aggregation. We then compute the standard errors for (8), and find that we reject at the 1% confidence level that the justified multipliers are above 0.915, 0.644, and 0.036, respectively. Overall, because demand does not predict future earnings, the price multipliers that can be justified by CF information are bounded and significantly smaller than the empirically estimated price multipliers that we later report in Section 3.

Clarifications and caveats. Two caveats are in order. First, we are not saying that demand *never* predicts CF. Rather, we believe that demand must be informative in some circumstances, and prior literature has indeed revealed examples of CF-informed trading. For instance, Hu, Pan, and Wang (2017) find that investors with an "early peek advantage" trade on their superior knowledge ahead of information releases. Our results show merely that, *on average*, our demand measure does not contain enough information to justify large price multipliers in a Kyle (1985) framework.

Second, data limitations prevent us from examining information about earnings more than five years out. In principle, it is possible that, even though demand does not predict earnings within five years, it strongly predicts longer-horizon earnings and thus justifies large information-based price effects. We do not consider this very likely because our demand measure captures aggressive trading by investors who are eager to trade quickly. Investors with access to long-horizon information do not have to trade in a rush and thus can reduce the use of aggressive orders. Having said this, not being able to study earnings information more than five years out is a limitation for our paper.

3 Estimating Price Multipliers at Different Levels

In this section, we estimate the price multipliers associated with different demand components. Throughout all specifications, we find that price multipliers are larger at more aggregate levels.

3.1 Estimating price multipliers

We estimate price multipliers using a panel regression of contemporaneous weekly returns on the decomposed demand components; the decomposition procedure was described in Section 2.3. The regression is

$$\begin{aligned}\text{Ret}_{i,t} = a + M^{3\times 3} \cdot \text{Demand}^{3\times 3}_{l1(i),t} + M^{6\times 6} \cdot \text{Demand}^{6\times 6}_{l2(i),t} \\ + M^{\text{Idio}} \cdot \text{Demand}^{\text{Idio}}_{i,t} + \text{Controls}_{i,t} + \epsilon_{i,t}\end{aligned} \tag{9}$$

where the controls include week fixed effects and a list of possible return predictors: four lags of weekly returns to capture short-term reversals as well as the characteristics included in the Fama-French six-factor model (Fama and French, 2018). We obtain the stock characteristics from Andrew Chen's website (Chen and Zimmermann, 2021) and standardize the characteristics into uniform distributions on [-0.5, 0.5] for each cross-section, following the practice in Kelly, Pruitt, and Su (2019). Standard errors are clustered by week and stock. It is worth noting that, by including week fixed effects, we exclude market-level variation and focus on price effects in the cross-section.

The full-sample regression result is reported in column (1) of Table 3. The price multiplier associated with the most aggregate 3×3 style demand is 6.976, implying that investing one dollar in a 3×3 portfolio increases the market valuation of that portfolio by 6.976 dollars. The multipliers associated with the less aggregate 6×6 style and idiosyncratic components are 3.121 and 1.732, respectively. For column (1d), we compute the pairwise differences between multipliers and find statistically significant differences. Overall, the results indicate

Table 3. Price multipliers estimated at weekly frequency

We report panel regression estimates of the impact of weekly demand on contemporaneous stock returns. Demand is decomposed into a 3×3 size-BM style component, a 6×6 size-BM style component, and an idiosyncratic component, as described in section 2.3. Column (1) reports full sample estimates and column (1d) reports the pairwise differences between coefficients (e.g. $\text{Demand}_{i,t}^{3\times 3} - \text{Demand}_{i,t}^{6\times 6}$). Columns (3) through (6) report estimates for four equal-length sub-periods. Columns (2) and (2d) are similar to columns (1) and (1d) except that we estimate the regression with demand decomposed using bootstrapped placebo stock characteristics. In all regressions we control for week fixed effects, the previous four weeks of stock returns, and the Fama-French six-factor characteristics. Standard errors are clustered by week and stock and reported in parentheses. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable: weekly stock return ($\text{Ret}_{i,t}$)							
	Actual data		Bootstrapped data		Actual data, sub-samples			
	Estimates	Differences	Estimates	Differences	1993-1999	2000-2007	2008-2015	2016-2022
	(1)	(1d)	(2)	(2d)	(3)	(4)	(5)	(6)
$\text{Demand}_{i,t}^{3\times 3}$	6.976***	3.856***	1.705***	0.039	7.910***	8.681***	4.738***	5.372***
	(0.450)	(0.475)	(0.086)	(0.101)	(0.487)	(0.856)	(0.473)	(1.696)
$\text{Demand}_{i,t}^{6\times 6}$	3.121***	1.388***	1.666***	−0.090	3.271***	3.952***	1.889***	2.297***
	(0.153)	(0.157)	(0.053)	(0.064)	(0.150)	(0.361)	(0.170)	(0.380)
$\text{Demand}_{i,t}^{\text{Idio}}$	1.732***		1.756***		2.278***	1.925***	0.786***	1.113***
	(0.036)		(0.037)		(0.040)	(0.070)	(0.052)	(0.080)
Week FE	Y		Y		Y	Y	Y	Y
Other controls	Y		Y		Y	Y	Y	Y
Obs	3,922,489		3,922,489		1,000,760	1,169,628	977,086	775,015
Within R^2	0.038		0.036		0.086	0.047	0.011	0.009

that price multipliers are monotonically larger for demand at more aggregate levels.

One may be worried that changes in trading practices and market liquidity over time may impact our inferences. For instance, as investors become more proficient at splitting orders and avoiding aggressive trading, one may be reasonably concerned that our demand measure becomes less useful at capturing trading in later periods. For instance, Bogousslavsky and Muravyev (2023) document that the amount of trading in daily closing auctions has been increasing over the past decade, reaching 7.5% of daily volume in 2018, and the Lee-Ready algorithm is not designed for signing the closing auction. There are also notable changes in market rules such as the NYSE decimalization in early 2001, the introduction of NYSE auto-quoting in 2003 (Hendershott, Jones, and Menkveld, 2011), and the rise of high-frequency trading since the introduction of Reg NMS in 2005.

To alleviate this concern, we estimate price multipliers by sub-samples and find that our general conclusion is robust over time. In columns (3) through (6), we divide the sample into four equal-length periods and re-estimate the regressions using each sub-sample. Within each sub-period, we find the same pattern that price multipliers are larger at more aggregate levels.

Models such as Duffie (2010) and Gabaix and Koijen (2022) generally predict that price impact primarily comes from surprise, rather than expected, demand. We verify this in Appendix F.1: if we decompose demand into an expected component — spanned by lagged demand and returns — and a residual surprise component, we find that price effects almost entirely arise from the surprise component. However, at the weekly frequency, most demand variation is unexpected and we find that less than 10% of demand variation can be predicted at any level of aggregation. Therefore we stick to our simpler specification and do not use decomposed demand in our main specification.

3.2 Robustness to alternative demand decomposition

One may be concerned that our result is specific to our demand decomposition procedure. To investigate this concern, we conduct three sets of additional exercises. First, we verify that the results are not mechanically caused by the decomposition procedure. Because our style-level demand is computed by averaging stock-level demand, one may worry that our results arise from the mechanical process of averaging. Specifically, if stock-level demand is measured with idiosyncratic error, such error will tend to wash out when averaging over a style portfolio. Because measurement errors in independent variables lead to attenuation biases, the idiosyncratic-level multipliers may be biased down while style-level multipliers are not, and that explains our findings.

This alternative hypothesis is easy to test. If the mere act of averaging caused our results, then using randomly shuffled stock characteristics to conduct demand decomposition should lead to similar findings. In each cross-section, we randomly match each stock to a different stock without replacement and use the matched stock's size and BM characteristics to re-conduct demand decomposition. In other words, we still apply the same demand decomposition procedure but use bootstrapped —and thus meaningless—stock characteristics. The results are shown in column (2) of Table 3 and exhibit no multiplier differences across different levels of aggregation, indicating that our results are not mechanical consequences of the demand decomposition procedure. This finding also indicates that price multipliers are only large when considering portfolios of stocks that are *meaningfully related*, and portfolios of unrelated stocks do not command higher multipliers.

Second, one may be concerned that our results are specific to the use of size and BM stock characteristics. To alleviate this concern, we also conduct alternative demand decompositions using pairwise combinations of commonly used stock characteristics: size, BM, investment, operating profitability, and momentum. This yields a total of ten distinct specifications and we estimate price multipliers for all specifications. In Appendix C.2.1 we show that, throughout all these specifications, price multipliers at more aggregate levels are larger, and the differences across different levels are always statistically significant at the 1% level. Therefore, our finding is not specific to using size or BM characteristics.

Third, we also check the robustness of our results by using flexible combinations of additional stock characteristics to decompose stock demand. For this purpose, we use the instrumented principal component analysis (IPCA) approach developed in Kelly et al. (2019). They develop an algorithm that takes many stock characteristics as inputs and conducts a flexible principal component (PC) decomposition in which the PC portfolio weights are transformations of stock characteristics. They identify the PC weights by maximizing the ability of the model to explain return covariance. The key benefit of this approach is that it builds a model with a bottom-up approach and thus we do not need to take a stance regarding which characteristics should be used.

We use the code and the 36 stock characteristics provided in the original Kelly et al. (2019) paper to conduct the decomposition. We categorize the 36 PCs identified by the model into two equal-sized groups to form the “more aggregate” and “less aggregate” demand components; the residual demand forms the idiosyncratic component. We then estimate the price multipliers associated with demand at these levels. The results are qualitatively similar to our main findings: more aggregate demand is associated with larger price multipliers. The methodology and the results are shown in greater detail in Appendix C.2.2.

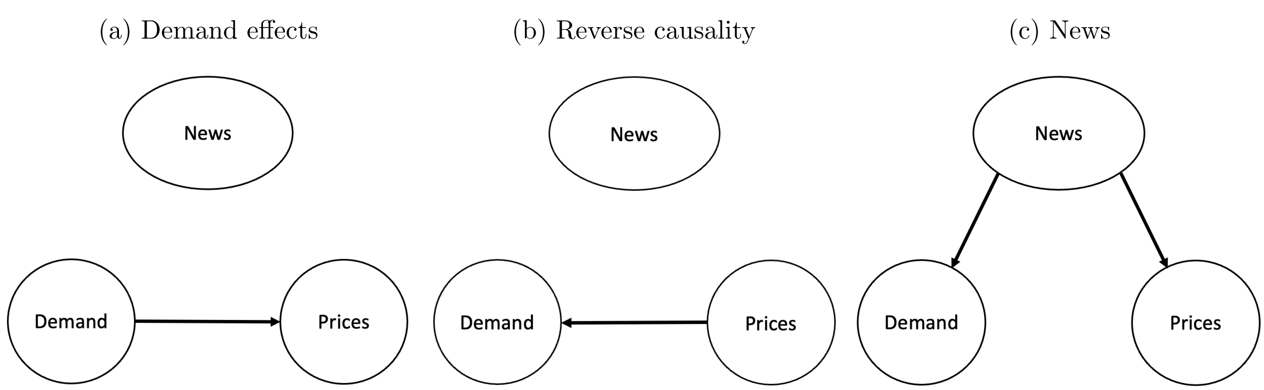
4 Alternative Hypotheses

We hypothesize that our price multiplier estimates primarily reflect demand effects, and the mechanism is illustrated in Panel (a) of Figure 2: demand directly causes price movements. There are two main alternative hypotheses that are illustrated in the other two Panels. In Panel (b), causality is reversed and demand follows prices. In Panel (c), demand and prices are both driven by news, but there is no direct relationship between demand and prices.

We discuss alternative hypotheses in this section. Section 4.1 studies the first alternative hypotheses while sections 4.2 and 4.3 examine the second.

Figure 2. Graphical illustration of alternative hypotheses

We illustrate different hypotheses for our findings using causal diagrams in which nodes represent variables and arrows represent causal relationships. In plot (a), demand drives price impact. In plot (b), the causality is reversed and price movements cause demand changes. In plot (c), both demand and prices are driven by news.



4.1 Reverse causality

It is natural to worry about reverse causality. Prior work has documented various instances in which demand chases past returns such as in the case of mutual fund flows (e.g. Chevalier and Ellison, 1997) and retail trading (e.g. Barber and Odean, 2008; Boehmer, Jones, Zhang, and Zhang, 2021). Therefore, one may be concerned that, because we use weekly data, our results may reflect return-chasing behavior at higher frequencies. To offer a concrete example, suppose that daily demand chases the previous day's return. This will show up as a spurious positive price multiplier at the weekly frequency even if demand has no direct impact on returns. Further, if return-chasing is somehow more prevalent for more aggregate demand components, this may explain our finding of larger price multipliers at more aggregate levels.

To examine this alternative hypothesis, we re-compute our demand measure at higher frequencies and re-estimate price multipliers. If our results arise from return chasing, when we use higher frequency data, the results should increasingly reflect genuinely contemporaneous relationships and thus the price multipliers should shrink. In contrast, if demand does create price impacts, then we expect the price multipliers to be generally unchanged, or even stronger, as higher-frequency results do not include short-term price reversions.

We re-estimate price multipliers at daily frequencies by subsamples and report the results in columns (1) through (4) of Panel A in Table 4. In all sub-samples, we observe patterns consistent with the weekly results: price multipliers are larger at more aggregate levels. Columns (1d) through (4d) report the pairwise differences in multiplier estimates and find all of them to be statistically significant at the 1% level.

We then tackle the concern of possible intraday return-chasing by re- estimating results using hourly and 10-minute demand and returns. The intraday return data is computed using the latest mid-point price quotes in each time interval. The intraday demand data is computed using the Lee-Ready algorithm implementation in Holden and Jacobsen (2014) and, to be consistent with our daily and weekly data, we also pre-process it by subtracting a stock-specific twenty-day moving average over the same intraday time period. We then decompose demand into three components using the methodology in section 2.3 and aggregate data at hourly and ten-minute intervals. For hourly data, the first "hour" is defined as 9:30:00 to 10:00:00 while the others are full hours such as 10:00:01 to 11:00:00. We do not use the overnight component of returns from the previous day's market close to this day's market open.

The intraday results reported in Panel B of Table 4 are also consistent with the earlier finding that multipliers are larger at more aggregate levels. While not reported for brevity, in all these regressions, the estimated price multipliers across different levels are statistically different from each other at the 1% level of statistical significance. To avoid bid-ask bounce-type effects, the returns in the independent variable are computed using both the current and the subsequent period. For instance, in the 10-minute regressions, demand over the period [9:30, 9:40] is used to explain returns over the period [9:40, 9:50]. Overall, our main findings hold up as we use increasingly higher-frequency data for the estimations, which is inconsistent with the return-chasing mechanism.

The intraday multipliers in Table 4 appear higher than the weekly ones in Table 3. Appendix F.2 shows that this is primarily explained by autocorrelated intraday demand. Intuitively, if each 1 unit of demand indicates the arrival of another $\beta - 1$ units in the subsequent periods (thus a total of β units overall), then the prima facie multiplier estimate needs to be adjusted by dividing by β. This doesn't affect the weekly estimate because weekly demand does not exhibit reversal or continuation (Figure A.3), but intraday demand is positively autocorrelated and requires adjustment. Once this adjustment is performed, price multiplier estimates are similar across all observational frequencies.

4.2 Controlling for directional news measures

The main remaining concern is whether our results are explained by information, as illustrated in Panel (c) of Figure 2. In section 2.4, we do not find evidence that demand

Table 4. Price multipliers estimated at higher frequencies

This table is similar to the last four columns of Table 3 except that the regressions are estimated using higher frequency data. We estimate panel regressions of stock returns on demand which is decomposed into three components. Columns (1) through (4) of Panel A report regression coefficients based on daily data and columns (1d) through (4d) report coefficient differences between different levels. In Panel B, we report results based on hourly data in columns (1) through (4) and results based on 10-minute data in columns (5) through (8). In all regressions, we control for time fixed effect corresponding to the time periods for each regression, as well as the list of return controls in Table 3. Standard errors are clustered by time and stock and reported in parentheses. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Panel A: daily frequency

	Estimates			
	1993-1999	2000-2007	2008-2015	2016-2022
	(1)	(2)	(3)	(4)
$\text{Demand}^{3\times3}_{i,t}$	9.080***	8.783***	5.171***	9.775***
	(0.288)	(0.713)	(0.710)	(0.784)
$\text{Demand}^{6\times6}_{i,t}$	3.591***	4.049***	2.119***	2.795***
	(0.093)	(0.207)	(0.147)	(0.206)
$\text{Demand}^{\text{Idio}}_{i,t}$	2.509***	2.229***	0.990***	1.267***
	(0.033)	(0.052)	(0.047)	(0.070)
Time FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
Obs	4,803,170	5,616,629	4,707,851	3,725,170
Within R^2	0.066	0.037	0.011	0.008

	Coefficient differences			
	1993-1999	2000-2007	2008-2015	2016-2022
	(1d)	(2d)	(3d)	(4d)
$\text{Demand}^{3\times3}_{i,t}$ - $\text{Demand}^{6\times6}_{i,t}$	5.488***	4.734***	3.052***	6.980***
	(0.303)	(0.743)	(0.725)	(0.811)
$\text{Demand}^{6\times6}_{i,t}$ - $\text{Demand}^{\text{Idio}}_{i,t}$	1.083***	1.820***	1.129***	1.528***
	(0.099)	(0.214)	(0.154)	(0.218)

Panel B: intraday frequencies

	Hourly frequency			
	1993-1999	2000-2007	2008-2015	2016-2022
	(1)	(2)	(3)	(4)
$\text{Demand}^{3\times3}_{i,t}$	11.293***	16.031***	13.923***	13.515***
	(0.276)	(0.494)	(0.447)	(1.028)
$\text{Demand}^{6\times6}_{i,t}$	4.756***	6.577***	6.803***	6.224***
	(0.079)	(0.195)	(0.173)	(0.194)
$\text{Demand}^{\text{Idio}}_{i,t}$	3.554***	3.637***	4.734***	4.197***
	(0.034)	(0.043)	(0.059)	(0.068)
Time FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
Obs	19,520,456	34,782,809	30,065,032	25,732,187
Within R^2	0.036	0.015	0.015	0.020

	10 minute frequency			
	1993-1999	2000-2007	2008-2015	2016-2022
	(5)	(6)	(7)	(8)
$\text{Demand}^{3\times3}_{i,t}$	10.197***	16.777***	13.959***	15.703***
	(0.157)	(0.434)	(0.450)	(1.250)
$\text{Demand}^{6\times6}_{i,t}$	4.931***	6.065***	7.468***	7.645***
	(0.053)	(0.111)	(0.153)	(0.205)
$\text{Demand}^{\text{Idio}}_{i,t}$	4.038***	3.973***	5.600***	5.101***
	(0.040)	(0.043)	(0.062)	(0.081)
Time FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
Obs	48,125,259	161,228,374	161,113,820	137,757,837
Within R^2	0.034	0.010	0.015	0.024

forecasts future cash flow news over the next five years. However, it is still possible that demand is correlated with *contemporaneous* news, and we examine this possibility here and in section 4.3.

In this section, we control for two commonly studied news measures with directional implications on stock returns. We briefly introduce the measures here and place data construction details in Appendix D.1. The first measure is IBES standardized unexpected earnings (SUE), which captures earnings surprises in each quarterly release. To also capture news during non-earnings periods, our second measure is the Ravenpack media coverage sentiment score called ESS (event sentiment score). Because there tends to be more news coverage of larger firms and during the later portion of the sample period, we standardize the ESS measure to zero mean and unit variance by firm-year. For comparability, the SUE variable is also standardized to have unit variance. Both measures have been shown to explain contemporaneous stock returns (Livnat and Mendenhall, 2006; Boudoukh, Feldman, Kogan, and Richardson, 2019), a fact that we confirm in our sample and report in Appendix Figure D.8.

We now consider whether price-multiplier estimates are affected when controlling for these news measures. We merge our main sample with news data and the resulting sample run from 2000 through 2021.[5] For column (1) in Table 5 we estimate price multipliers in the merged sample and find the familiar pattern whereby multipliers decline from higher to lower levels of aggregation. For column (2), we estimate the same regression using SUE instead of demand as the main independent variable, and we decompose SUE into three components following the same decomposition procedure as in section 2.3. We do the same for column (3) using ESS. The results indicate that both SUE and ESS carry statistically significant explanatory power over returns.

For columns (4) through (6), we control for SUE and ESS when estimating demand-based price multipliers. The multiplier estimates are almost entirely unaffected, indicating that the demand effects are orthogonal to our firm-level news measures. In Panel B of Appendix Table D.11, we verify that demand has effectively zero correlation with either SUE or ESS. Overall, we do not find evidence that our price multipliers can be explained by contemporaneous news releases reflected in these two commonly used measures.

[5]The data limitation reflects Ravenpack's database, which starts in 2000. Further, the authors' institutions have purchased updates only until 2021, so the last year of the sample period, 2022, is not covered. We do not believe that adding one more year of data will materially change our conclusions.

Table 5. Controlling for news measures in multiplier estimation

This table is analogous to Table 3 but with additional control variables. We merge our weekly sample with IBES standardized unexpected earnings (SUE) and the Ravenpack event sentiment score (ESS), two stock-level news measures. SUE and ESS are both standardized, converted to percentages, and decomposed into three components following the procedure described in section 2.3. For columns (1) through (3) we include one type of regressor at a time. For columns (4) through (6) we add SUE and ESS controls to examine whether demand-based price multipliers are affected. In all regressions we control for week fixed effects, the previous four weeks of stock returns, and the Fama-French six-factor characteristics. Standard errors are clustered by week and stock and reported in parentheses. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable: weekly stock return ($\text{Ret}_{i,t}$)					
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Demand}_{i,t}^{3\times3}$	6.218***			6.219***	6.209***	6.219***
	(0.446)			(0.445)	(0.444)	(0.444)
$\text{Demand}_{i,t}^{6\times6}$	2.573***			2.569***	2.563***	2.564***
	(0.155)			(0.154)	(0.154)	(0.154)
$\text{Demand}_{i,t}^{\text{Idio}}$	1.192***			1.196***	1.179***	1.185***
	(0.041)			(0.040)	(0.040)	(0.040)
$\text{SUE}_{i,t}^{3\times3}$		0.558***		0.619***		0.530***
		(0.178)		(0.167)		(0.176)
$\text{SUE}_{i,t}^{6\times6}$		0.609***		0.613***		0.450***
		(0.056)		(0.054)		(0.053)
$\text{SUE}_{i,t}^{\text{Idio}}$		0.694***		0.695***		0.502***
		(0.012)		(0.012)		(0.012)
$\text{ESS}_{i,t}^{3\times3}$			0.402***		0.323**	0.160
			(0.138)		(0.129)	(0.136)
$\text{ESS}_{i,t}^{6\times6}$			0.638***		0.629***	0.484***
			(0.055)		(0.053)	(0.052)
$\text{ESS}_{i,t}^{\text{Idio}}$			0.809***		0.802***	0.628***
			(0.014)		(0.014)	(0.014)
Week FE	Y	Y	Y	Y	Y	Y
Other controls	Y	Y	Y	Y	Y	Y
Obs	2,230,709	2,230,709	2,230,709	2,230,709	2,230,709	2,230,709
Within R^2	0.019	0.016	0.020	0.035	0.038	0.045

4.3 Controlling for non-directional news indicators

In section 4.2 we show that demand-based price effects are not explained by either of two commonly studied news measures. While this is comforting, those two measures capture only a subset of value-relevant news. There are additional indicators of news that, unlike SUE or ESS, cannot be converted into a directional score but does measure the *volume* of news that arrives in each period. In other words, while the directional news measures discussed in section 4.2 can explain *returns*, other non-directional news indicators can explain *volatility*, and we would like to control for them as well.

In this section, we sort the sample into subsamples with more or less news and estimate price multipliers on them separately. If our results are driven by demand effects, we should expect our findings to be robust across subsamples. According to the information-based alternative hypothesis, price multipliers should be smaller in subsamples with less news. We now describe the news indicators we use.

Stock-level news indicators. Our first indicator is the number of IBES analyst updates in each stock-week. It is widely documented that analysts respond to new information by updating their forecasts (e.g. Conrad, Cornell, Landsman, and Rountree, 2006). Savor (2012) find that large price movements associated with analyst reports tend to exhibit continuation while those not associated with analyst updates tend to exhibit reversion, which is consistent with the idea that analyst updates convey value-relevant information.

Andersen, Bollerslev, Diebold, and Labys (2003) advocate using high-frequency returns to compute realized volatility (rvol). We follow them and compute rvol for stock-weeks using hourly stock returns computed from changes in mid quotes. Because rvol makes use of all price movements, it is a rather inclusive measure of information arrival. However, we cannot directly use *realized* contemporaneous rvol because it is endogenous, and comparing price multipliers on rvol-sorted subsamples generates spurious results (Appendix D.3). Luckily, rvol is highly persistent in the cross-section, so using lagged rvol can explain a high fraction of realized rvol and proxy for the *expected* (rather than *realized*) rvol, so we use lagged rvol as our second news indicator. Finally, we also continue to use the Ravenpack ESS measure but, instead of using it as a directional measure, we take its absolute value so it also becomes an indicator of the volume of arriving news.

Macro-level news indicators. In addition to controlling for firm-level indicators, we also control for three macro-level news indicators. The first measure follows Savor and Wilson

(2013) and Savor and Wilson (2014) to consider weeks with Federal Reserve announcements, non-farm employment reports, and inflation (PPI/CPI) releases as periods when macroeconomic news is released. Savor and Wilson (2013) find that over 60% of U.S. stock-market expected returns were realized in periods when these three releases occurred, indicating that they are important news events. We obtained data up to 2011 from the publicly shared files of Savor and Wilson (2014) and extended these data to cover our sample period using news release dates from the Federal Reserve website and the St. Louis Fed Archival FRED website. Overall, according to this indicator, macroeconomic news is released in 54% of sample weeks.

While the three abovementioned releases may be the most important, we also want to capture other macroeconomic releases. To do so, we choose as our second measure the Aruoba-Diebold-Scotti (ADS) Real Time Business Conditions Index, which combines more macroeconomic data sources and is designed to track real business conditions at daily frequency (Aruoba, Diebold, and Scotti, 2009). We download the latest version from the Philadelphia Federal Reserve website and use the absolute value of weekly index changes to measure the volume of macroeconomic news in that week.

While our first two measures capture information from formal economic releases, with our third measure we aim to capture additional news beyond formal releases. For this purpose, we download news articles related to the United States from Ravenpack's global macroeconomic database (RPA 1.0). In addition to economic releases, this source also contains a wide range of news and reporting about business, the environment, politics, and social issues. Following Ravenpack's recommendation, we screen out the lower-quality news source "MRVR," require the news-relevance score to be at least 90, and require the "novelty" score to be at least 90, enabling us to avoid re-capturing stale news. Like the stock-level Ravenpack data, each article is associated with an event sentiment score (ESS) $\in [-1, 1]$, and we use the absolute value of the sum of all articles' ESSs to measure the aggregate macroeconomic news in a given week.

Price multiplier estimates by news indicator-sorted subsamples. We merge all these news indicators with our data and use the joint sample in subsequent analyses. Some news indicators vary mechanically across firms and over time. For instance, the IBES analyst updates indicator is usually higher for large-cap stocks as a result of more analyst coverage and rises over time as the number of analysts increases. The same issue impacts the stock- and macro-level Ravenpack news indicators, which also rise over time as additional news sources are collected by Ravenpack. Therefore, we transform all stock-level (and macro-level) news indicators using ranks into uniform distributions over [0,1] for each stock-year (and year). In Appendix D.2 we show that periods with higher news indicators tend to exhibit higher return volatility, suggesting that these indicators are indeed related to news.

We now sort the sample using these news indicators into subsamples and estimate price multipliers on each indicator. To study the effects of stock-level news, we first remove all earnings weeks, as those are clearly associated with cash-flow news, and then sort the resulting sample into $2 \times 2 \times 2 = 8$ subsamples based on whether each of the three indicators—IBES updates, Ravenpack ESS, and rvol—are above or below their respective median values. Because all news indicators are standardized by stock-year, the subsamples are balanced over time and across stocks.

The multiplier estimates by subsamples are reported in Panel A of Table 6. For instance, in column (3) we report results for the subsample with below-median ("low") news according to the number of IBES analyst updates, above-median ("high") news according to Ravenpack ESS, and below-median ("low") news according to rvol. While we observe some variation in multiplier estimates across subsamples, within each subsample the ordering of price multipliers is the same, and the pairwise differences (e.g. $\text{Demand}^{3\times3} - \text{Demand}^{6\times6}$) are always statistically significant at the 5% level. Most importantly, we do not see evidence for the information-based alternative hypothesis. Even in the column (1) subsample where all news indicators are low, the price multipliers are not significantly smaller than in other subsamples.

Table 6. Price multipliers in subsamples with more or less news

We estimate price multipliers in subsamples sorted by news indicators into low and high news periods. The regression methodology is the same as in Table 3. For Panel A, we sort the sample into $2 \times 2 \times 2$ subsamples based on whether three stock-level news indicators are above or below median. For instance, in column (3) we report results for the subsample with below-median ("low") news according to the number of IBES analyst updates, above-median ("high") news according to Ravenpack —ESS—, and below-median ("low") news according to realized volatility. For Panel B, in a similar way, we sort the sample into $2 \times 2 \times 2$ subsamples using three macroeconomic news indicators: whether the week contains one of the macroeconomic news releases studied in Savor and Wilson (2013), the absolute value of changes in the Aruoba-Diebold-Scotti (ADS) Business Conditions Index, and Ravenpack global macroeconomic news ESS. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Dependent variable: weekly stock return ($\text{Ret}_{i,t}$)

Panel A: sorting by stock-level news

IBES	Low (L)				High (H)			
Ravenpack	L		H		L		H	
Reazlied vol	L	H	L	H	L	H	L	H
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Demand}^{3\times3}_{i,t}$	5.584***	6.424***	5.877***	6.474***	6.276***	6.861***	5.936***	6.188***
	(0.364)	(0.620)	(0.418)	(0.633)	(0.431)	(0.684)	(0.458)	(0.757)
$\text{Demand}^{6\times6}_{i,t}$	2.680***	2.883***	2.597***	2.900***	2.798***	3.003***	2.368***	2.603***
	(0.174)	(0.235)	(0.215)	(0.266)	(0.192)	(0.257)	(0.225)	(0.317)
$\text{Demand}^{\text{Idio}}_{i,t}$	1.516***	1.585***	1.412***	1.411***	1.584***	1.672***	1.235***	1.130***
	(0.049)	(0.055)	(0.053)	(0.060)	(0.049)	(0.059)	(0.055)	(0.063)
Week FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Obs	297,001	270,519	244,886	234,121	235,626	243,381	269,014	298,506
Within R^2	0.038	0.036	0.033	0.029	0.039	0.034	0.022	0.016

Panel B: sorting by macro news

Savor-Wilson	Low (L)				High (H)			
ADS	L		H		L		H	
Ravenpack	L	H	L	H	L	H	L	H
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Demand}^{3\times3}_{i,t}$	6.140***	7.754***	8.976***	4.600***	5.621***	5.767***	7.528***	6.593***
	(0.816)	(1.239)	(1.490)	(1.177)	(1.699)	(0.976)	(1.373)	(0.814)
$\text{Demand}^{6\times6}_{i,t}$	2.508***	3.558***	3.912***	2.204***	2.921***	2.450***	3.329***	3.186***
	(0.363)	(0.486)	(0.588)	(0.377)	(0.700)	(0.368)	(0.629)	(0.386)
$\text{Demand}^{\text{Idio}}_{i,t}$	1.444***	1.585***	1.816***	1.276***	1.442***	1.472***	1.501***	1.583***
	(0.101)	(0.117)	(0.138)	(0.114)	(0.130)	(0.106)	(0.112)	(0.117)
Week FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Obs	315,556	272,043	283,009	273,231	347,323	319,375	308,409	389,647
Within R^2	0.029	0.036	0.041	0.025	0.031	0.029	0.035	0.031

We then conduct a similar exercise by examining $2 \times 2 \times 2$ subsamples formed using the three macroeconomic news indicators and report the results in Panel B of Table 6.

The conclusion is the same: Even in the subsample with low news according to all three indicators, we still find larger price multipliers at higher levels of aggregation. Overall, the results do not support the information-based alternative hypothesis.

5 Theoretical interpretations

The prior literature has proposed several types of mechanisms that might drive price impacts and. In this section, we compare them against our findings. In section 5.1 we examine the *qualitative* features of those mechanisms and argue that risk-based explanations are the most promising: They generate larger price multipliers at higher levels of aggregation, while other mechanisms do not. When examining the *quantitative* predictions, we find and report in section 5.2 that standard risk-based models tend to predict overly small idiosyncratic price multipliers. Section 5.3 ends with a tentative discussion of possible resolutions of this discrepancy.

5.1 Qualitative fit of mechanisms

We summarize the qualitative predictions associated with the main theoretical mechanisms in Figure 3. The black dots mark our empirical estimates of price multipliers as reported in Table 3 and the whiskers represent 95% confidence intervals. As discussed, we find price multipliers to be larger at higher levels of aggregation. The three lines illustrate the predictions associated with three general classes of theoretical models. Of these, only risk-based models are aligned with our findings in predicting higher aggregate-level price multipliers. We discuss these models briefly below and present analytical details in Appendix E.

Risk-based explanations. Risk-based equilibrium asset-pricing models (e.g. Petajisto, 2009; Kozak et al., 2018) fit our findings the best. In these models, price movements can

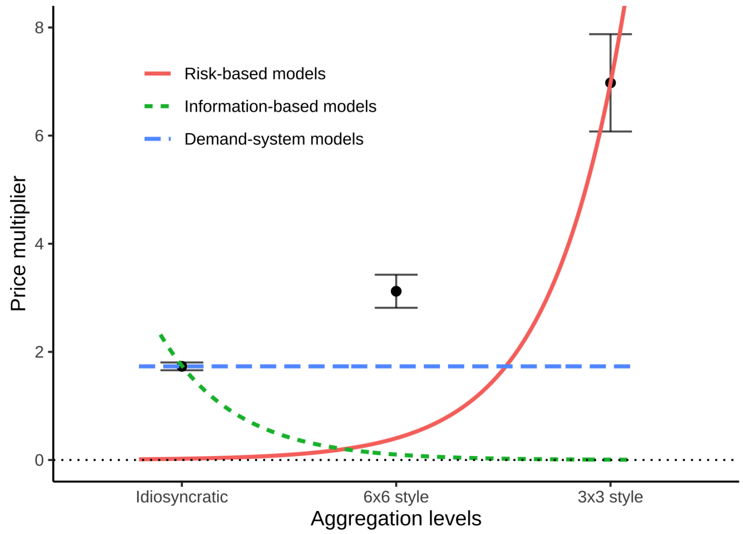


Figure 3. An illustration of the predictions of several mechanisms

We illustrate qualitatively the predictions of various price-impact mechanisms and compare them against our empirical findings. The black dots mark our empirical estimates of price multipliers at multiple levels of aggregation reported in Table 3 and the whiskers represent 95% confidence intervals. The red line illustrates predictions associated with risk-based models (e.g. Petajisto, 2009; Kozak et al., 2018). The green dashed line illustrates information-based models, which generally predict higher multipliers at the idiosyncratic level (e.g. Subrahmanyam, 1991). The blue dashed line illustrates demand-system models a la Koijen and Yogo (2019) which predict, to a first-order approximation, equal price multipliers at all levels.

be driven by exogenous supply shocks (Petajisto, 2009) or demand shocks from the trading by sentiment investors (Kozak et al., 2018), and the price effects are determined in equilibrium by the absorption of these shocks by risk-averse investors. Because liquidity-providing investors choose optimal portfolios, they are much more willing to trade against idiosyncratic-level mispricing, which is less risky. In contrast, investors are more reluctant to trade against price dislocations at higher levels of aggregation, as doing so exposes them to less diversifiable risk. Such investor behavior naturally leads to larger price multipliers at higher levels of aggregation. This is a general prediction generated by all risk-based models. Later in section 5.2 we make this concrete using a simple model.

Information-based explanations. Models such as those in Kyle (1985) and Glosten and Milgrom (1985) predict long-lasting trading-induced price impacts that reflect private information about cash flows. The empirical tests we present in sections 2.4 and 4.2 do not find any relationship between demand and cash flows. Therefore, these models are less likely to be applicable to our findings, but it remains possible that our cash-flow measures are not sufficiently comprehensive, so we consider this mechanism once again here.

It is worth noting that information-based models tend to predict smaller price multipliers at higher levels of aggregation (e.g. Subrahmanyam, 1991; Gorton and Pennacchi, 1993). This is because they assume that there is more asymmetric information about cash flows at the stock level that washes out at higher levels of aggregation. This assumption is also often made in empirical papers. For instance, Chordia et al. (2002) (p113) states that, "for the aggregate market, asymmetric information seems unlikely...". However, our empirical tests robustly find larger multipliers at higher levels of aggregation. Therefore, information-based explanations are inconsistent with our findings.

Logit demand-system models. Koijen and Yogo (2019) (henceforth KY) and many follow-up papers use characteristics-based logit functions to model investor demand. Their approach has been widely applied to quantifying demand-based price effects in stocks (e.g. Koijen et al., Forthcoming; Van der Beck, 2022; Haddad, Huebner, and Loualiche, 2022; Tamoni, Sokolinski, and Li, 2023), corporate bonds (Bretscher, Schmid, Sen, and Sharma, 2022), and government bonds (Koijen, Koulischer, Nguyen, and Yogo, 2021; Jansen, 2023).

To a first-order approximation, these models predict equal price multipliers at all levels of aggregation. We discuss the KY model briefly here and present further details in Appendix E.1. In KY, each investor i chooses her log portfolio weight in security n to be

$$\log(w_i(n)) \approx \beta_{0,i} \cdot \log P(n) + \sum_{k=1}^{K-1} \beta_{k,i} \cdot x_k(n) + \beta_{K,i} + \log \epsilon_i(n), \tag{10}$$

where the characteristics include the stock price $P(n)$ and a set of non-price characteristics $\{x_k(n)\}_{k=1}^{K-1}$. $\{\beta_{k,i}\}_{k=0}^{K}$ are preference parameters, and $\epsilon_i(n)$ is a latent demand term that captures demand variation that is unrelated to the characteristics.

In the baseline KY model, investor wealth A_i is exogenous and the investor holds $Q_i(n) =$

$A_i \cdot w_i(n)/P(n)$ shares of stock n. Therefore, her stock-level demand elasticity is:

$$
\begin{aligned}
-\frac{d\log Q_i(n)}{d\log P(n)} &= 1 - \frac{d\log(A_i \cdot w_i(n))}{d\log P(n)} \\
&= 1 - \frac{d\log(w_i(n))}{d\log P(n)} \\
&\overset{(10)}{\approx} 1 - \beta_{0,i}.
\end{aligned} \tag{11}
$$

How does this compare with style-level demand elasticities? In other words, if the prices of all stocks in the same style portfolio changed at the same time, how would the investor's demand respond? In this baseline model, style-level demand elasticities are approximately the same as stock-level demand elasticities. To see this, note that, in equation (10), demand for stock n depends only on its *own* price, not the prices of other stocks. Therefore, as price multipliers equal the reciprocal of the demand elasticities in equilibrium models, KY predicts, to a first-order approximation, equal price multipliers at all levels. Appendix E.1 shows that model extensions that endogenize wealth effects or incorporate investor heterogeneity also do not differ much in this regard.

5.2 Quantitative fit of mechanisms

Section 5.1 shows that risk-based models provide the best qualitative fit to our findings. However, upon closer examination, there is still a quantitative discrepancy. Risk-based models predict vanishingly small price multipliers at idiosyncratic levels, and this is because they assume that investors can effectively hedge away non-aggregate risks. More concretely, they generally predict that a style portfolio with N stocks should command price multipliers that are $\mathcal{O}(N)$ that of the idiosyncratic level. Therefore, when moving from higher to lower levels of aggregation, price multipliers should decline very quickly, as illustrated by the blue line in Figure 3: If we calibrate a model to match the 3×3 style-level multiplier, it will necessarily underestimate the idiosyncratic-level multiplier by a significant amount.

We now explain the scaling of multipliers across levels using a simple model. Appendix

E.2 provides derivation details.

Model set-up. Consider a one-period model where investors trade at time 0 and payoffs are realized at time 1. The risk-free rate is normalized to zero. There are $i = 1, ..., N$ stocks and their shares outstanding $\{Z_i\}_{i=1}^N$ are normalized to 1. Each stock belongs to one of $s = 1, ..., S$ equal-sized style portfolios with $N_s = N/S$ stocks in each and, to simplify the notation, we assume N_s is an integer. In later calibration, we will think of these styles as corresponding to the 3×3 size/BM-sorted styles in our empirical exercises.

Each stock i has a terminal payoff at time 1 given by

$$X_i = a_i + F_m + F_{s(i)} + \epsilon_i, \tag{12}$$

where $s(i) \in \{1, 2, ..., S\}$ is the style to which i belongs and $a_i = E(X_i) > 0$ is a constant. The other terms are mean-zero shocks at the market-, style-, and idiosyncratic levels, respectively:

$$F_m \sim \mathcal{N}(0, \sigma^2_{\text{market}}) \tag{13}$$

$$F_s \sim \mathcal{N}(0, \sigma^2_{\text{style}}) \quad \forall s = 1, ..., S \tag{14}$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2_{\text{idio}}) \quad \forall i = 1, ..., N. \tag{15}$$

We assume that all shocks occur independently of each other. Prices $\{P_i\}_{i=1}^N$ are set by trading at time 0. There is a unit mass of mean-variance investors who can be aggregated into a representative investor with CARA utility and a coefficient of absolute risk aversion γ. The representative investor optimization problem is

$$\max_{\{D_i\}_{i=1}^N} E\left[-\exp\left(-\gamma W\right)\right] \quad \text{where} \quad W = W_0 + \sum_{i=1}^N D_i \cdot (X_i - P_i), \tag{16}$$

where $\{D_i\}_{i=1}^N$ are the number of shares demanded, W_0 is the initial wealth, and W is the terminal wealth. Equating investor demand with supply gives the equilibrium expected returns,

$$
\begin{aligned}
E(R_i) &= a_i - P_i \\
&= \gamma \left[\underbrace{\left(N \cdot \sigma^2_{\text{market}}\right) \cdot Z^{\text{market}}}_{\text{market-level risk}} + \underbrace{\left(N_s \cdot \sigma^2_{\text{style}}\right) \cdot Z^{\text{style}}_{s(i)}}_{\text{style-level risk}} + \underbrace{\sigma^2_{\text{idio}} \cdot Z_i}_{\text{idiosyncratic risk}} \right],
\end{aligned} \tag{17}
$$

where $Z^{\text{market}} \equiv \frac{1}{N}\sum_{i=1}^{N} Z_i$ and $Z^{\text{style}}_s \equiv \frac{1}{N_s}\sum_{i\in s} Z_i$ are the average market-level and style-level share supplies, respectively. These variables are equal to one because we have normalized the stock-level share supply to 1. Equation (17) shows that, in decomposing the risk premium for stock i, the importance of supply at different levels is proportional to the *overall* amount of risk at each level.

The specification of equation (17) is akin to the regressions we use to estimate price multipliers (equation (9)). In a one-period model, changes in expected returns are exactly opposite to the impact on current prices. Further, because the supply variables in equation (17) are all normalized to unity, the coefficients with which they are associated are exactly price multipliers. Concretely, equation (17) implies, for instance, that a 1% change in stock-level supply Z_i should create a price impact of approximately $1\% \times \gamma\sigma^2_{\text{idio}}$. Therefore, the ratio of style- and stock-level price multipliers is given by the ratio between the total amount of risks:

$$
\frac{M^{\text{style}}}{M^{\text{idio}}} = \frac{\gamma N_s \sigma^2_{\text{style}}}{\gamma \sigma^2_{\text{idio}}} = \frac{N_s \sigma^2_{\text{style}}}{\sigma^2_{\text{idio}}}. \tag{18}
$$

This ratio scales with N_s, the number of stocks in a style. As long as σ_{style} and σ_{idio} are not overly different from one another, this model implies that style-level multipliers are significantly larger. As we report in Appendix E.2, we calibrate this model to the data, taking the 3×3 size/BM-sorted portfolios as styles in the model, and the calibration predicts that $\frac{M^{\text{style}}}{M^{\text{idio}}} \approx 19$. This is significantly higher than the ratio of $6.976/1.732 \approx 4$ in estimated data (Table 3).

Two points are worth clarifying. First, the discrepancy is not about the average size of price multipliers but about scaling across levels. In the language of the model, the risk-aversion parameter γ governs the average size of price multipliers, but it cancels out when taking the ratio in equation (18). If the average multiplier size is off, one can always change assumptions about risk aversion or arbitrageur capital to fit the data, but these do not affect the scaling. Second, while we derive this prediction using a specific CARA-normal model, the prediction of a significantly larger style-level price multiplier is common to risk-based models because, in these models, price multipliers are determined by how risk premiums change in response to supply shocks. To an optimizing investor, the idiosyncratic-level supply is considered much less risky so it must command a very small risk premium and, as a consequence, a small price multiplier.

5.3 Possible extensions to the mechanism

As discussed so far in this section, risk-based models provide the best explanation for our findings, but additional "fixes" are needed to explain why idiosyncratic-level price multipliers are larger than theory predicts. We outline two possible explanations below.

1. **Limited diversification.** Asset-pricing models often assume that investors view stocks as good substitutes and form optimal portfolios that diversify away from idiosyncratic risks. As a consequence, stock-level risk commands little, if any, risk premium. However, if investors do not diversify as much as frictionless models assume, stock-level risk will command a higher risk premium and stock-level price multipliers will rise.

   There are several possible reasons for limited diversification. Some investors may hold concentrated portfolios. Even for investors who hold many stocks, if they do not substitute between stocks as flexibly as standard theory predicts, their behavior can still generate higher stock-level multipliers. Imperfect substitution may arise from using heuristics in portfolio construction, such as the $1/N$ or risk-parity strategies

(DeMiguel, Garlappi, and Uppal, 2009; Asness, Frazzini, and Pedersen, 2012). Even for fully optimizing mean-variance investors, Davis, Kargar, and Li (2023) show that they may still exhibit lower stock-level demand elasticity as a result of limited cross-sectional spanning between stocks. Asset-pricing models often predict that stocks are almost perfectly spanned by combinations of other stocks, but empirical studies find that factor models often fail to adequately span stock-level expected returns (e.g. Lopez-Lira and Roussanov, 2022). As a consequence, investors may rationally view stocks as imperfect substitutes.

2. **Behavioral perception of information.** In this paper, we do not find evidence that our demand measure is related to cash-flow information. However, market participants may not know this and assume—or at least fear—that demand carries information. As discussed in section 5.1, asymmetric information is often seen as being more relevant at the stock level, so this mechanism primarily increases stock-level price multipliers.

We think both explanations are plausible; it is also possible that both mechanisms are at play and may even interact with each other. We leave the task of separating these explanations for future work.

6 Summary

The price multiplier is a key parameter for understanding how investor demand shapes asset prices. Using Lee-Ready order-flow imbalance in the stock market, this paper finds that price multipliers form a continuum along which they are larger at higher levels of aggregation. Our results are largely consistent with risk-based mechanisms where investors are reluctant to take on undiversifiable risk, and they are not consistent with explanations based on information about cash flows (Kyle, 1985).

Empirically, we find that price multipliers associated with long-short–style portfolios are several times larger than the idiosyncratic-level multiplier. This feature is not incorporated into the demand system approach pioneered by Koijen and Yogo (2019), which predicts that style-level multipliers are, to a first-order approximation, the same as stock-level multipliers. Therefore, applying the model to style- or factor-levels directly may underestimate the magnitude of demand-induced price effects (e.g. Van der Beck, 2022; Tamoni et al., 2023). Similarly, existing studies of the profit-generating capacity of factor (anomaly) strategies also assume that stock-level transaction costs apply to factor-level trading (e.g. Ratcliffe et al., 2017), which can lead researchers to overestimate profit capacities.

We are hopeful that our demand measure can become another price instrument in the empiricists' toolbox, operating alongside the commonly used stock index changes (Shleifer, 1986; Chang et al., 2015) and flow-induced trading instruments (Lou, 2012). Our demand measure exhibits ample variation at many levels of aggregation, which allows it to have strong first-stage explanatory power for prices. We also show that our measure does not contain information about earnings over the next five years. Having said this, we need to caution that our demand measure captures only aggressive trades, so the price multipliers we estimate may not apply to passively executed trades.

References

Abarbanell, J. S., B. J. Bushee, and J. Smith Raedy. 2003. Institutional investor preferences and price pressure: The case of corporate spin-offs. *The Journal of Business* 76:233–261.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2003. Modeling and forecasting realized volatility. *Econometrica* 71:579–625.

Aruoba, S. B., F. X. Diebold, and C. Scotti. 2009. Real-time measurement of business conditions. *Journal of Business & Economic Statistics* 27:417–427.

Asness, C. S., A. Frazzini, and L. H. Pedersen. 2012. Leverage aversion and risk parity. *Financial Analysts Journal* 68:47–59.

Barber, B. M., and T. Odean. 2007. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

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* 21:785–818.

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

Ben-David, I., F. Franzoni, and R. Moussawi. 2018. Do ETFs increase volatility? *Journal of Finance* 73:2471–2535.

Ben-David, I., F. Franzoni, R. Moussawi, and J. Sedunov. 2021a. The granular nature of large institutional investors. *Management Science* pp. Advanced access published April 30, 2021, 10.1287/mnsc.2020.3808.

Ben-David, I., J. Li, A. Rossi, and Y. Song. 2021b. What do mutual fund investors really care about? *Review of Financial Studies* .

Ben-David, I., J. Li, A. Rossi, and Y. Song. 2022. Ratings-Driven Demand and Systematic Price Fluctuations. *The Review of Financial Studies* .

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang. 2021. Tracking retail investor activity. *The Journal of Finance* 76:2249–2305.

Bogousslavsky, V., and D. Muravyev. 2023. Who trades at the close? Implications for price discovery and liquidity. *Journal of Financial Markets* p. 100852.

Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson. 2019. Information, trading, and volatility: Evidence from firm-specific news. *The Review of Financial Studies* 32:992–1033.

Boyer, B., and L. Zheng. 2009. Investor flows and stock market returns. *Journal of Empirical Finance* 16:87–100.

Breen, W. J., L. S. Hodrick, and R. A. Korajczyk. 2002. Predicting equity liquidity. *Management Science* 48:470–483.

Bretscher, L., L. Schmid, I. Sen, and V. Sharma. 2022. Institutional corporate bond pricing. *Swiss Finance Institute Research Paper* .

Brown, D. C., S. Davies, and M. Ringgenberg. 2021. ETF arbitrage and return predictability. *Review of Finance* 25:937–972.

Brunnermeier, M., E. Farhi, R. S. Koijen, A. Krishnamurthy, S. C. Ludvigson, H. Lustig, S. Nagel, and M. Piazzesi. 2021. Perspectives on the Future of Asset Pricing. *The review of financial studies* .

Budish, E., P. Cramton, and J. Shim. 2015. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics* 130:1547–1621.

Campbell, J. Y., and R. J. Shiller. 1988. The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies* 1:195–228.

Chakrabarty, B., P. C. Moulton, and A. Shkilko. 2012. Short sales, long sales, and the Lee–Ready trade classification algorithm revisited. *Journal of Financial Markets* 15:467–491.

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

Chen, A. Y., and T. Zimmermann. 2021. Open Source Cross Sectional Asset Pricing. *Critical Finance Review* .

Chen, H. 2022. Cash induced demand. *Journal of Financial and Quantitative Analysis* pp. 1–26.

Chen, W., S. Liang, and D. Shi. 2023. What Drives Stock Prices in a Bubble? *Working paper* .

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

Chordia, T., R. Roll, and A. Subrahmanyam. 2002. Order imbalance, liquidity, and market returns. *Journal of Financial economics* 65:111–130.

Chordia, T., and A. Subrahmanyam. 2004. Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics* 72:485–518.

Chu, Q. C., W.-l. G. Hsieh, and Y. Tse. 1999. Price discovery on the S&P 500 index markets: An analysis of spot index, index futures, and SPDRs. *International Review of Financial Analysis* 8:21–34.

Conrad, J., B. Cornell, W. R. Landsman, and B. R. Rountree. 2006. How do analyst recommendations respond to major news? *Journal of Financial and Quantitative Analysis* 41:25–49.

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

Darmouni, O., K. Siani, and K. Xiao. 2022. Nonbank fragility in credit markets: Evidence from a two-layer asset demand system. *Working paper* .

Davis, C., M. Kargar, and J. Li. 2023. Why is Asset Demand Inelastic? .

DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The review of Financial studies* 22:1915–1953.

Deuskar, P., and T. C. Johnson. 2011. Market liquidity and flow-driven risk. *The Review of Financial Studies* 24:721–753.

Driscoll, J. C., and A. C. Kraay. 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics* 80:549–560.

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

Easterwood, J. C., and S. R. Nutt. 1999. Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *The Journal of Finance* 54:1777–1797.

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

Fama, E. F., and K. R. French. 2018. Choosing factors. *Journal of financial economics* 128:234–252.

Frazzini, A., R. Israel, and T. J. Moskowitz. 2020. Trading costs of asset pricing anomalies. *Working paper* .

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

Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting characteristics nonparametrically. *The Review of Financial Studies* 33:2326–2377.

Froot, K., and M. Teo. 2008. Style investing and institutional investors. *Journal of Financial and Quantitative Analysis* 43:883–906.

Gabaix, X., and R. S. Koijen. 2022. In search of the origins of financial fluctuations: The inelastic markets hypothesis. *Working paper* .

Glosten, L. R., and P. R. Milgrom. 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics* 14:71–100.

Gorton, G. B., and G. G. Pennacchi. 1993. Security Baskets and Index-Linked Securities. *Journal of Business* pp. 1–27.

Haddad, V., P. Huebner, and E. Loualiche. 2022. How Competitive is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing. *Working paper* .

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. *Journal of Finance* 41:815–829.

Hartzmark, S. M., and D. H. Solomon. 2023. Market-Wide Predictable price pressure. *Working paper* .

Hasbrouck, J. 2003. Intraday price formation in US equity index markets. *The Journal of Finance* 58:2375–2400.

Hasbrouck, J., and D. J. Seppi. 2001. Common factors in prices, order flows, and liquidity. *Journal of financial Economics* 59:383–411.

Hendershott, T., C. M. Jones, and A. J. Menkveld. 2011. Does algorithmic trading improve liquidity? *Journal of Finance* 66:1–33.

Hirshleifer, D., S. S. Lim, and S. H. Teoh. 2009. Driven to distraction: Extraneous events and underreaction to earnings news. *The journal of finance* 64:2289–2325.

Holden, C. W., and S. Jacobsen. 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *The Journal of Finance* 69:1747–1785.

Hu, G. X., J. Pan, and J. Wang. 2017. Early peek advantage? Efficient price discovery with tiered information disclosure. *Journal of Financial Economics* 126:399–421.

Huang, S., Y. Song, and H. Xiang. 2022. Noise trading and asset pricing factors. *Working paper* .

Jansen, K. A. 2023. Long-term investors, demand shifts, and yields. *Working paper* .

Jegadeesh, N., and J. Livnat. 2006. Revenue surprises and stock returns. *Journal of Accounting and Economics* 41:147–171.

Jordà, Ò. 2005. Estimation and inference of impulse responses by local projections. *American economic review* 95:161–182.

Keim, D. B., and A. Madhavan. 1995. Anatomy of the trading process empirical evidence on the behavior of institutional traders. *Journal of Financial Economics* 37:371–398.

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

Kelly, B. T., S. Pruitt, and Y. Su. 2020. Instrumented principal component analysis. *Available at SSRN 2983919* .

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* 140:1–20.

Koijen, R. S., R. J. Richmond, and M. Yogo. Forthcoming. Which investors matter for equity valuations and expected returns? *Review of Economic Studies* .

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

Korajczyk, R. A., and R. Sadka. 2004. Are momentum profits robust to trading costs? *The Journal of Finance* 59:1039–1082.

Kozak, S., S. Nagel, and S. Santosh. 2018. Interpreting factor models. *Journal of Finance* 73:1183–1223.

Kumar, A., and C. M. Lee. 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61:2451–2486.

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

Lee, C. M., and M. J. Ready. 1991. Inferring trade direction from intraday data. *The Journal of Finance* 46:733–746.

Lewellen, J., et al. 2015. The Cross-section of Expected Stock Returns. *Critical Finance Review* 4:1–44.

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

Li, J. J., N. D. Pearson, and Q. Zhang. 2021. Impact of demand shocks on the stock market: Evidence from Chinese IPOs. *Neil D. and Zhang, Qi, Impact of Demand Shocks on the Stock Market: Evidence from Chinese IPOs (December 16, 2021)* .

Liu, C., and B. Wang. 2021. Demand curves for stocks slope down in the long run: evidence from the Chinese split-share structure reform. *Critical Finance Review forthcoming* .

Livnat, J., and R. R. Mendenhall. 2006. Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of accounting research* 44:177–205.

Lopez-Lira, A., and N. L. Roussanov. 2022. Do Common Factors Really Explain the Cross-Section of Stock Returns? *Working paper* .

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

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

Menkveld, A. J., et al. 2023. Non-standard errors. *Journal of Finance* .

Novy-Marx, R., and M. Velikov. 2016. A taxonomy of anomalies and their trading costs. *The Review of Financial Studies* 29:104–147.

Parker, J. A., A. Schoar, and Y. Sun. 2020. Retail financial innovation and stock market dynamics: The case of target date funds. Tech. rep., National Bureau of Economic Research.

Patel, N., and I. Welch. 2017. Extended stock returns in response to S&P 500 index changes. *Review of Asset Pricing Studies* .

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

Peng, C., and C. Wang. 2021. Factor Demand and Factor Returns. *Working paper* .

Petajisto, A. 2009. Why do demand curves for stocks slope down? *Journal of Financial and Quantitative Analysis* 44:1013–1044.

Ratcliffe, R., P. Miranda, and A. Ang. 2017. Capacity of smart beta strategies from a transaction cost perspective. *The Journal of Index Investing* 8:39–50.

Savor, P., and M. Wilson. 2013. How much do investors care about macroeconomic risk? Evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis* 48:343–375.

Savor, P., and M. Wilson. 2014. Asset pricing: A tale of two days. *Journal of Financial Economics* 113:171–201.

Savor, P. G. 2012. Stock returns after major price shocks: The impact of information. *Journal of financial Economics* 106:635–659.

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

Schmickler, S. N. 2021. Payout-Induced Trading, Asset Demand Elasticities, and Market Feedback Effects .

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

Subrahmanyam, A. 1991. A theory of trading in stock index futures. *The Review of Financial Studies* 4:17–51.

Tamoni, A., S. Sokolinski, and Y. Li. 2023. Which Investors Drive Anomaly Returns and How? *Available at SSRN 4242745* .

Teo, M., and S.-J. Woo. 2004. Style effects in the cross-section of stock returns. *Journal of Financial Economics* 74:367–398.

Van der Beck, P. 2022. Flow-driven ESG returns. *Working paper* .

Vayanos, D., and P. Woolley. 2013. An institutional theory of momentum and reversal. *The Review of Financial Studies* 26:1087–1145.

Vuolteenaho, T. 2002. What Drives Firm-Level Stock Returns? *Journal of Finance* 57:233–264.

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

APPENDIX

This appendix presents supplementary analytical and empirical results. Appendix A provides details about demand measure construction and Appendix B studies its properties. Appendix C presents additional results on estimating price multipliers. Appendix D presents additional details on testing the information-based alternative hypothesis. Appendix E provides derivations on theoretical predictions about price impact.

A Measuring demand

This appendix section presents further details about measuring demand using Lee-Ready order flow imbalances.

A.1 Pre-processing demand data

Lee-Ready cannot sign some trades. As discussed in section 2.1, when the Lee-Ready algorithm is able to sign trades as buys or sells, the error rate is low when we aggregate data at slower frequencies. However, some trades in the earlier part of the trading day can be hard to sign due to not having earlier prices to compare to, and our demand measure will end up missing this trade.

How large is the missing data problem? On average, trades in the first five (thirty) minute of each trading day accounts for around 3.5% (13.2%) of total volume and, over the full sample, 20.2% (9.6%) of those trades cannot be signed by Lee-Ready. The problem primarily affects the sample up to 2008 and almost all trades can be signed after that. This is not solved by using the DTAQ data instead of MTAQ before 2008 as similar issues exist in the DTAQ data.[6] Later in this section, Figure A.2 provides more detail about the distribution of unsigned trades.

[6]We thank Jun Wu at WRDS for helping us investigate this.

Missing a subset of trades that have no *systematic* imbalance would not substantially bias our results: we would simply have a classical measurement error in demand. However, Lou, Polk, and Skouras (2019) find that trades around the open tend to be correlated and are often in the opposite direction with trades during the day. Specifically, they argue that trades around the open disproportionally reflect trades by retail investors and tend to exert more buying pressure on smaller companies than larger ones, while trades during the day — which our demand measure fully captures — tend to exhibit the opposite pattern.

This can create a spurious autocorrelation in style-level demand. Concretely, suppose as Lou et al. (2019) indicated, traders around the open tend to be net buyers of small-cap stocks and net sellers of large-cap stocks, and traders during the day do the opposite. Even if overall trading is balanced, because the Lee-Ready algorithm misses some of the trades around the open, it will detect persistent positive (negative) demand pressures in large-cap (small-cap) stocks.

We find evidence of this bias and illustrate it in Figure A.1. Panel (a) plots the average weekly demand for stocks sorted by size into quintiles. As expected, for the period up to 2008, large-cap stocks such as those in quintile five systematically have higher demand than small-cap stocks. In the post-2008 sample where we do not suffer from the missing data problem, the difference becomes muted.

A pre-processing step to reduce the bias. To alleviate this bias, we pre-process the stock-level daily demand data by subtracting its own moving average in the past 20 trading days (approximately a month):

$$\text{Demand}_{i,t} = \text{RawDemand}_{i,t} - \frac{1}{20}\sum_{d=1}^{20} \text{RawDemand}_{i,t-d} \tag{A1}$$

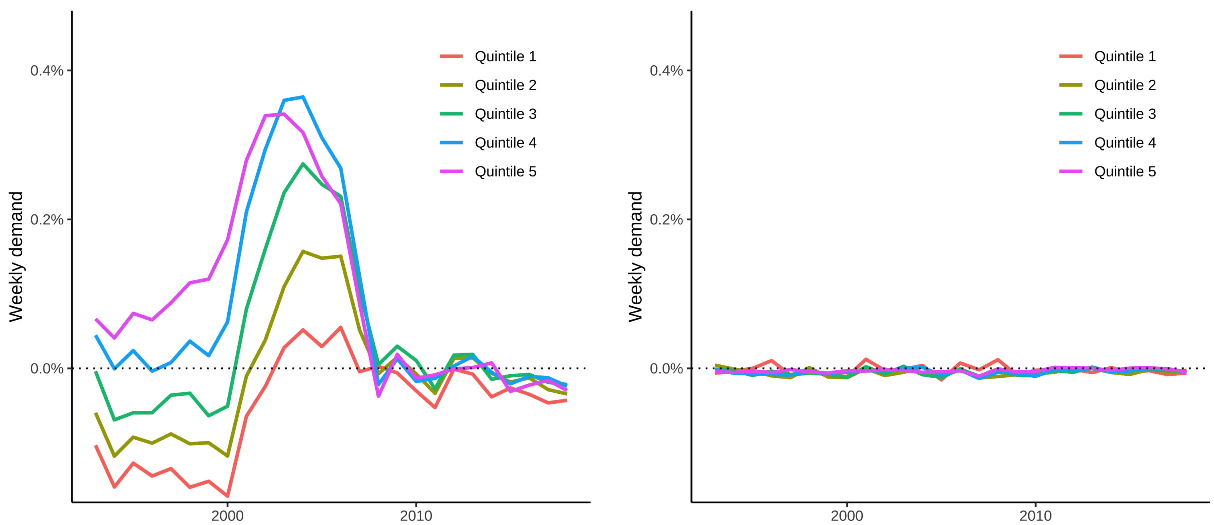
where $\text{RawDemand}_{i,t}$ is the raw daily order flow imbalance downloaded from WRDS Intraday Indicators divided by lagged shares outstanding. The idea is, if a stock suffers from the missing data problem, then it will tend to exhibit a slow-moving imbalance which we then

Figure A.1. Pre-processing demand data: before versus after

We pre-process the daily stock demand data from WRDS Intraday indicators by subtracting a 20-day moving average. This Figure plots the effect of this pre-processing step. We sort stocks cross-sectionally into quintiles by lagged market capitalization and plot the average weekly demand by year and quintile. Panel (a) plots the results before pre-processing and (b) plots the results after pre-processing.

(a) Before pre-processing

(b) After pre-processing



remove using this procedure. The results after this pre-processing step are plotted in Panel (b) of Figure A.1. The style-level biases in demand have been removed and there is no longer a spurious market cap-related demand persistence. In the main paper, our daily and weekly demand data are always based on this pre-processed version. When studying intraday demand data in section 4.1, the pre-processing step subtracts a twenty-day moving average of raw demand for the same intraday time period. For instance, consider demand during the hour of 10:00 to 11:00. We pre-process by subtracting the average demand during the same hour in the previous twenty trading days.

The pre-processing step *only* impacts the result in Appendix A.2 where we examine demand persistence. If we do not pre-process the data, as discussed earlier, we would spuriously find a higher autocorrelation for the 3×3 style-level demand than idiosyncratic-level demand.[7] We need to note that all our other results, and especially the main finding that price multipliers are larger at more aggregate levels, are not qualitatively affected by this pre-processing step. Our results are also similar when using longer or shorter look-back windows in pre-processing.

The rest of this section provides further background details about the reasoning behind our data pre-processing.

More details about trades that Lee-Ready does not sign. How much trading is there around the open and how many does Lee-Ready fail to sign? To gauge this, we ran the Lee-Ready algorithm through the entire TAQ data and summarized the results by five-minute intervals.

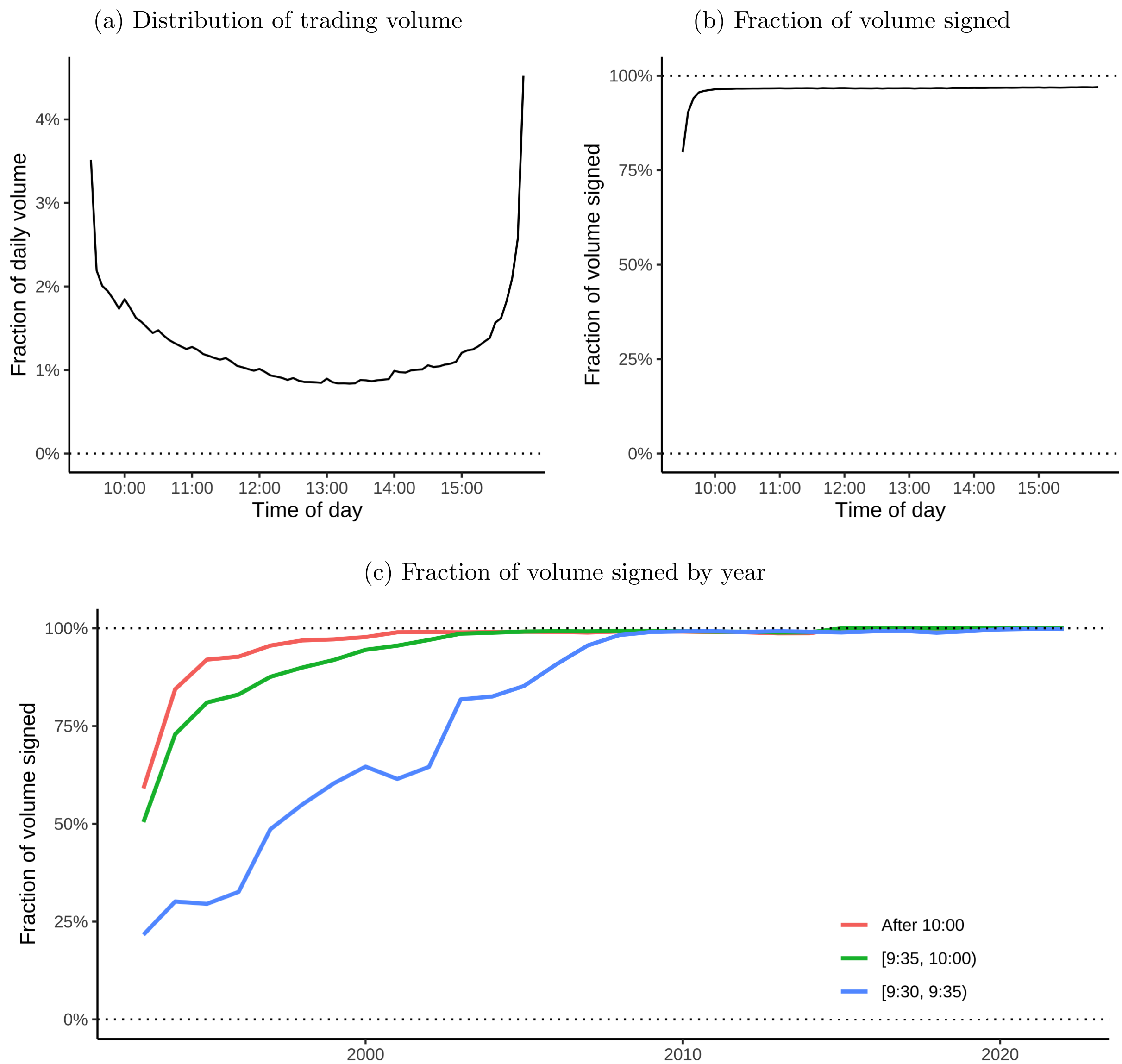
Panel (a) of Figure A.2 plots the distribution of dollar trading volume by five-minute intervals over the course of an average trading day over the full sample. On average, the first five (thirty) minutes of each trading day accounts for 3.5% (13.2%) of total trading. Panel (b) plots the fraction of volume by each five-minute interval that is successfully signed by Lee-Ready. For most periods, around 97% of trades can be successfully signed but the ratio dips around the open. Specifically, for the first five and thirty minutes, the average fraction signed is 79.8% and 90.4%, respectively.

The problem of missing trade signs is an issue in the sample up to 2008; almost all trades are signed since 2009. This is shown in Panel (c) where we plot the fraction of volume signed by year and intraday intervals. The blue, green, and red lines plot the fraction of signed trading volume for the first five minutes of each day, the next twenty-five minutes of each day, and the rest of the day, respectively. The problem of not being able to sign trades goes away after 2009.

[7]As shown in Panel (a) of Figure A.1, the measurement problem causes the large-cap portfolios to have higher demand than small-cap portfolios in the earlier sample but not in the later sample. Thus, in a full-sample panel regression with lags, we would end up with a large regression coefficient on the lags when studying style-level demand.

Figure A.2. Trades that Lee-Ready does not sign.

Panel (a) plots the fraction of dollar trading volume by five-minute intervals over an average trading day and Panel (b) plots the average fraction of trades by five-minute intervals that can be signed as buys or sells by the Lee-Ready algorithm. Panel (c) plots the fraction of volume signed by Lee-Ready for different intraday periods by year.



A.2 Statistical properties of the demand measure

Traditionally, microstructure studies tend to focus on order flow imbalances over short horizons which may give the wrong impression that our demand measure washes out at slower (e.g. weekly) frequencies. This is not true. In fact, we find little evidence that demand reverts at any level of aggregation, consistent with the findings based on an earlier 1988–1998 sample in Chordia and Subrahmanyam (2004). Persistence matters because in most asset pricing models, quickly mean-reverting demand should have little price effect.

Daily persistence. At each level of aggregation $l \in \{3\times 3 \text{ style}, 6\times 6 \text{ style}, \text{Idiosyncratic}\}$, we estimate panel regressions at daily frequency:

$$\text{Demand}^l_{i,t} = b_0 + \sum_{h=1}^{5} b_h \cdot \text{Demand}^l_{i,t-h} + \text{Controls}_{i,t} + \epsilon_{i,t}$$

and we cluster standard errors by day t and stock or style-portfolio i. Table A.1 reports regression results with day fixed effects in the first three columns and results with day and stock/portfolio fixed effects in the last three columns. Overall, we find that daily demand exhibits mild persistence for three to four days. The last row reports the implied half-life of demand in days based on the coefficient on the first lag (i.e. $\log(2)/\log(\widehat{\text{coef}})$). The implied half-life is around 0.3 to 0.35 days for all demand components.

Impulse response at longer horizons. We further examine the dynamics of demand at longer horizons using impulse responses estimated from weekly data. We follow the methodology in Jordà (2005) to estimate impulse responses using expanding windows. At each level of aggregation $l \in \{3 \times 3 \text{ style}, 6 \times 6 \text{ style}, \text{Idiosyncratic}\}$, we estimate panel regressions with expanding horizons of $h = 0, ..., 8$ weeks:

$$\sum_{s=0}^{h} \text{Demand}^l_{i,t+s} = a^l_h + b^l_h \cdot \text{Demand}^l_{i,t} + \tau^l_{h,t} + \epsilon^l_{i,h,t} \tag{A2}$$

where $\tau^l_{h,t}$ is the week fixed effect and we cluster standard errors by week t and stock or style-portfolio indicator i.

Panel (a) of Figure A.3 plots the resulting impulse responses traced out by estimated coefficients $\{\hat{b}^l_h\}_h$ estimated using the full sample. The results indicate that all demand

Table A.1. Daily demand persistence

This table uses panel regressions to estimate the dependence of daily demand on its own lags. As explained in section 2.3, stock-level demand is decomposed into three components at different levels of aggregation and we estimate separate panel regressions for each component. Columns (1) and (4) examine the most aggregate 3×3 size-BM style-level demand; columns (2) and (5) examine the less aggregate 6×6 style-level demand, and columns (3) and (6) examine idiosyncratic stock-level demand. All regressions control for day fixed effects and results in columns (4) to (6) also control for stock/style portfolio fixed effects. In all regressions, standard errors are clustered by day and stock/portfolio and reported in parentheses. The last row reports the implied half-life of demand based on the regression coefficient on $\text{Demand}_{i,t-1}$. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable: daily demand ($\text{Demand}_{i,t}$)					
	3×3 style	6×6 style	Idiosyncratic	3×3 style	6×6 style	Idiosyncratic
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Demand}_{i,t-1}$	0.137***	0.101***	0.094***	0.136***	0.101***	0.094***
	(0.021)	(0.004)	(0.002)	(0.020)	(0.004)	(0.002)
$\text{Demand}_{i,t-2}$	0.035***	0.032***	0.028***	0.034***	0.032***	0.028***
	(0.008)	(0.003)	(0.001)	(0.008)	(0.003)	(0.001)
$\text{Demand}_{i,t-3}$	0.025***	0.020***	0.013***	0.024**	0.019***	0.013***
	(0.009)	(0.003)	(0.001)	(0.009)	(0.003)	(0.001)
$\text{Demand}_{i,t-4}$	0.013	0.007**	0.005***	0.012	0.007**	0.005***
	(0.008)	(0.003)	(0.001)	(0.008)	(0.003)	(0.001)
$\text{Demand}_{i,t-5}$	−0.007	−0.004	−0.002**	−0.008	−0.004	−0.002***
	(0.014)	(0.003)	(0.001)	(0.014)	(0.003)	(0.001)
Day FE	Y	Y	Y	Y	Y	Y
Stock/portfolio FE	N	N	N	Y	Y	Y
Obs	67,878	271,457	18,524,398	67,878	271,457	18,524,398
Within R^2	0.023	0.013	0.011	0.022	0.013	0.010
Implied half-life (days)	0.349	0.302	0.293	0.348	0.302	0.293

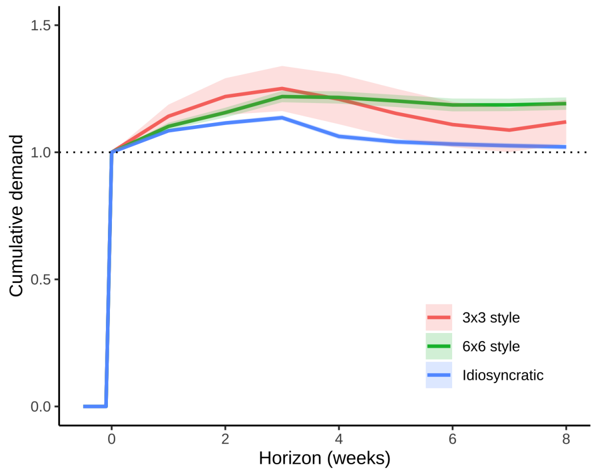
components exhibit slight momentum for two to three weeks but converge soon with no reversals. We then estimate impulse responses separately for three sub-periods and find qualitatively similar results and plot them in Panels (b) to (d) of Figure A.3. Overall, the results indicate that weekly demand appears to be permanent with no clear reversal or continuation.

As discussed in Appendix A.1, we have applied a pre-processing step of subtracting the previous 20-day moving average for each stock in the raw demand data. If not, the more aggregate style-levels will exhibit spurious autocorrelation which is removed by the pre-processing, as shown in Figure A.1. That said, the pre-processing step can also introduce mechanical negative autocorrelations which, luckily, is easy to estimate and subtract from

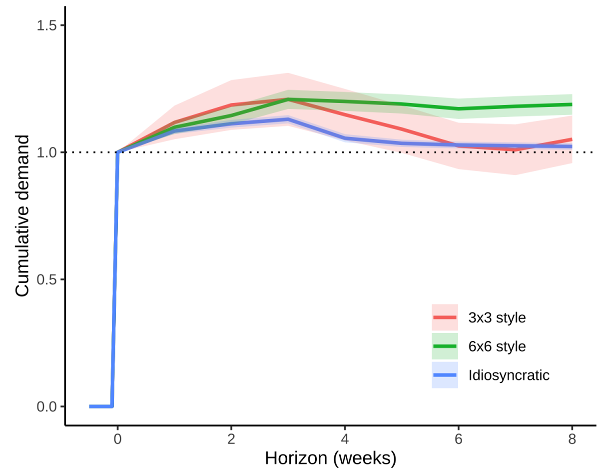
Figure A.3. Impulse responses of demand at longer horizons

This figure plots cumulative impulse responses of demand with 95% confidence bands. We estimate these impulse responses using panel regressions of the cumulative demand over weeks t to $t+h$ on week-t demand for $h = 0, ..., 8$. We estimate separate sets of regressions for demand at each decomposed component: the 3×3 style component, 6×6 style component, and the idiosyncratic level. The regressions control for week fixed effects and cluster standard errors by week and stock or style portfolio. Panel (a) plots results estimated over the full sample. Panels (b), (c), and (d) plot results estimated for different sub-periods. The horizontal dashed lines mark unity.

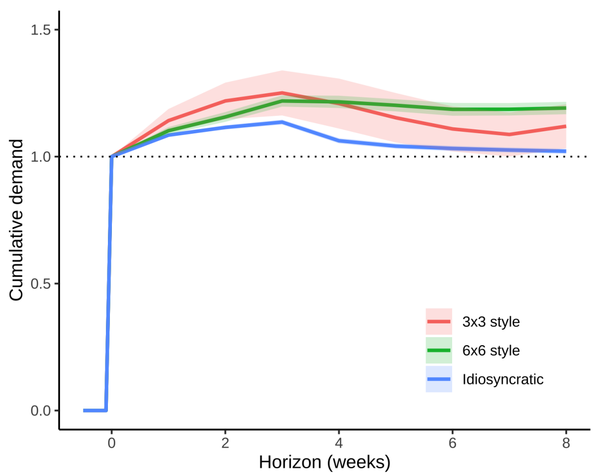
(a) Full sample



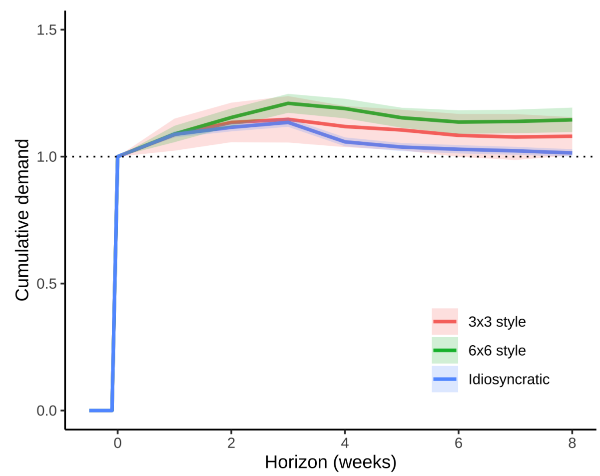
(b) Sub-sample: 1993 - 2002



(c) Sub-sample: 2003 - 2012



(d) Sub-sample: 2013 - 2022



the impulse responses. Specifically, we simulate i.i.d. zero-mean normally distributed raw demand data with a large sample size of $N = 10,000,000$, pre-process by subtracting the previous 20-day moving average as in Appendix A.1, and then estimate impulse responses using equation (A2). Because the simulated demand has no reversion by construction, any reversion in the resulting impulse response coefficients must arise mechanically from the pre-processing step. We then subtract this mechanical component from the impulse responses estimated from actual demand data.

B What does the demand measure capture?

This section presents additional details for examining the information content of the demand measure.

B.1 Demand around stock index changes

In this section, we document that our demand measure is correlated with the well-known non-CF trading by index funds around stock index changes. Since Shleifer (1986) and Harris and Gurel (1986), it has been well-known that when stocks are added to or deleted from the S&P 500 index, S&P-tracking index funds buy or sell those stocks to satisfy their investment mandates. We use these events to study whether our demand measure captures index change-induced trading.

Table B.2. Number of S&P 500 index-change events

In this table we report the number of S&P 500 index additions and deletions from 1993 through 2016 after merging these data with our dataset. The index events are obtained from Patel and Welch (2017) and then merged with our demand data.

	1993	1994	1995	1996	1997	1998	1999	2000	
Additions	7	14	28	22	29	43	35	51	
Deletions	9	13	31	15	25	40	37	43	
	2001	2002	2003	2004	2005	2006	2007	2008	
Additions	24	18	5	17	12	24	28	25	
Deletions	20	12	4	19	7	18	26	22	
	2009	2010	2011	2012	2013	2014	2015	2016	Total
Additions	22	11	7	7	11	8	17	16	481
Deletions	13	8	11	10	14	11	11	8	427

We obtain S&P index-change effective dates from the authors of Patel and Welch (2017) and merge these data with our demand data, which starts in 1993. The final sample covers 1993 through 2016 and contains 481 addition and 427 deletion events. The number of events per year is reported in Table B.2. We examine the behavior of demand for (t-10, t+10) days around each index event using a panel regression:

$$\begin{aligned}\text{Demand}_{i,t} = &\sum_{\tau=-10}^{10} \beta_\tau^{\text{addition}} \cdot \mathbf{I}_{\text{event time }\tau}^{\text{addition}} + \sum_{\tau=-10}^{10} \beta_\tau^{\text{deletion}} \cdot \mathbf{I}_{\text{event time }\tau}^{\text{deletion}} \\ &+ \text{Controls}_{i,t} + \epsilon_{i,t} \end{aligned} \tag{A3}$$

where the controls include date and stock fixed effects, and we cluster standard errors by date and stock. Dependent variables $\mathbf{I}_{\text{event time }\tau}^{\text{addition}}$ and $\mathbf{I}_{\text{event time }\tau}^{\text{deletion}}$ are indicators that equal one if day t is the τ^{th} day around an addition or a deletion event, respectively, for stock i. Therefore, coefficients $\beta_\tau^{\text{addition}}$ and $\beta_\tau^{\text{deletion}}$ estimate the abnormal demand associated with being $\tau = -10, ..., 10$ days around an addition or deletion event.

Panels (a) and (b) of Figure B.4 plot the results for addition and deletion events. Consistent with the idea that our demand measure captures trading by index funds, the measure positively (negatively) spikes exactly on index inclusion (deletion) days. The whiskers represent 95% confidence intervals and show that the demand changes are statistically significant. Overall, the results indicate that our measure indeed can capture CF-irrelevant shifts in investor demand.

B.2 Does demand predict future earnings?

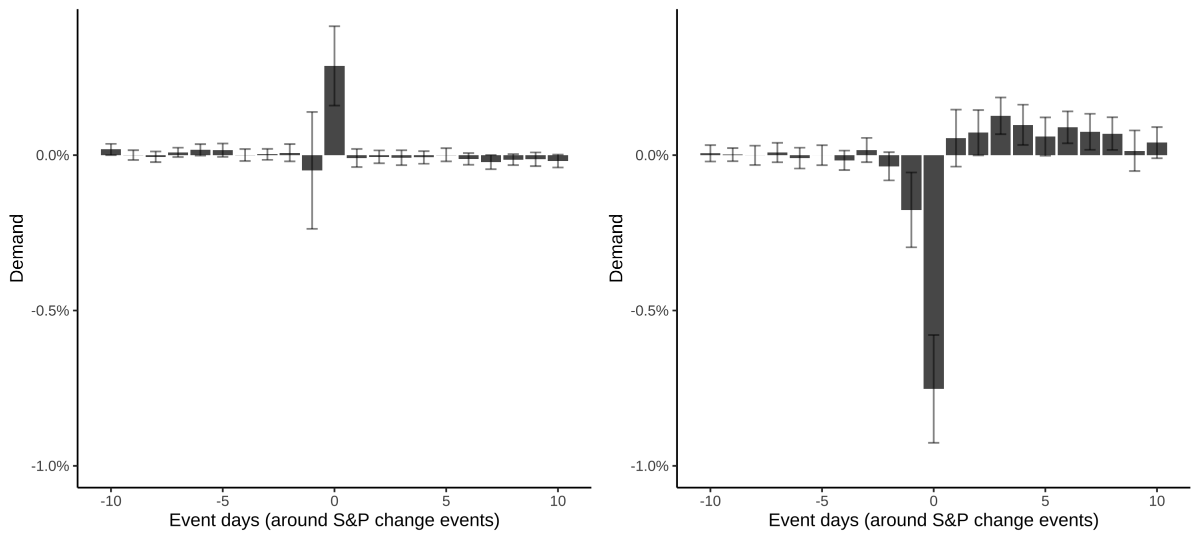
This section provides more details on the earnings forecast tests in section 2.4.

Data construction. We obtain the consensus mean analyst forecasts of earnings per share (EPS) and actual EPS realizations from IBES and merge them with our data. Because IBES is organized by monthly "statistical periods", we also align our demand and return data to this time grid. For instance, the first IBES monthly statistical period in 1993 starts on

Figure B.4. Demand around S&P 500 index-change dates

The two panels plot daily abnormal demand for (t-10, t+10) days around S&P 500 index addition and deletion events, respectively. The whiskers represent 95% confidence intervals; standard errors are clustered by date and stock. Estimation details are described in section B.1.

(a) Index addition events (b) Index deletion events



January 15 and ends on February 18. To align with the IBES time grid, we also construct a monthly demand variable by summing up daily demand during the same period. Stock returns over IBES periods are computed similarly.

Table B.3 provides summary statistics for the merging process. Panel A reports the fraction of firms for which we can find IBES data. For annual earnings, we can find forecasts of one- and two-year ahead annual earnings for at least two-thirds of the firms in all subperiods, but the coverage declines as we extend to three- to five-year ahead earnings. The lowest coverage appears in five-year-ahead EPS where we can find coverage for approximately one-seventh of the sample on average. In particular, coverage of long-horizon EPS declined particularly sharply towards the end of the sample because we also require having realized EPS releases. For instance, while we may find five-year-ahead EPS forecasts for firms in 2021, those EPS releases won't happen until 2025 or 2026. Quarterly EPS coverage is better and the average coverage is on the order of 70% to 80% over the full sample.

Table B.3. Summary statistics for IBES data coverage

We merge our demand and return data with IBES analyst forecasts and EPS realizations, and this Table reports the extent of IBES coverage. Panel A reports the fraction of companies that have IBES forecasts. Columns (1) to (5) report the fraction with one- to five-year ahead annual EPS forecasts and realizations, and columns (6) to (9) report the fraction with one- to four-quarter ahead quarterly EPS forecasts and realizations. The first six rows report coverage for five-year sub-periods and the last row reports the average over the full sample. Panel B reports the number of unique firms with IBES forecasts.

Panel A: fraction of sample with IBES forecasts

Period	Annual earnings forecasts					Quarterly earnings forecasts			
	1y ahead	2y	3y	4y	5y	1q ahead	2q	3q	4q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1993-1997	76.5%	70.1%	41.3%	15.2%	10.1%	75.0%	70.3%	66.6%	62.6%
1998-2002	77.2%	68.5%	41.5%	6.8%	2.4%	77.4%	74.5%	71.3%	67.9%
2003-2007	84.8%	79.0%	64.2%	27.9%	19.3%	85.1%	83.7%	81.7%	79.9%
2008-2012	86.3%	81.4%	73.4%	41.9%	30.2%	87.1%	86.1%	84.8%	83.3%
2013-2017	81.8%	77.0%	70.6%	40.9%	29.8%	82.3%	81.3%	80.2%	78.9%
2018-2022	72.5%	53.9%	36.0%	13.1%	3.9%	75.2%	74.2%	71.9%	58.0%
Full sample	79.6%	71.4%	53.3%	22.9%	14.9%	80.0%	77.9%	75.5%	71.3%

Panel B: number of unique firms with IBES forecasts

Period	Annual earnings forecasts					Quarterly earnings forecasts			
	1y ahead	2y	3y	4y	5y	1q ahead	2q	3q	4q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1993-1997	4,305	3,957	2,732	1,071	780	4,284	4,089	3,898	3,726
1998-2002	4,880	4,270	2,966	703	261	4,944	4,779	4,594	4,404
2003-2007	3,887	3,658	3,164	1,762	1,263	3,907	3,865	3,791	3,715
2008-2012	3,310	3,070	2,820	1,945	1,422	3,356	3,307	3,258	3,189
2013-2017	3,152	2,973	2,759	1,847	1,397	3,172	3,134	3,092	3,042
2018-2022	2,929	2,523	2,134	1,196	600	2,994	2,943	2,896	2,811
Full sample	9,170	8,115	6,358	3,749	2,731	9,334	9,060	8,786	8,508

B.2.1 Sub-sample results

Table 2 finds that demand has little forecasting power over future annual and quarterly EPS. This section shows that this is robust across sub-samples.

We split the sample into three equal-length periods and then estimate the relationship between demand and future EPS separately. The results are reported in the three panels of Table B.4. The specification and reporting format is the same as in Table 2: the first five columns examine the predictive power of demand over future annual EPS while the last four columns examine future quarterly EPS. Across all sub-samples, demand does not have sufficient EPS forecasting power to justify the price multipliers that we empirically estimate.

Table B.4. Using demand to predict earnings: sub-sample analysis

This table is the same as Table 2 except that we estimate by sub-samples. We split the sample into three equal-length periods and report the estimated results in three panels. In all panels, the first five columns report forecasting regressions on future annual EPS and the last four columns report forecasting regressions on future quarterly EPS. To save space, regression coefficients on lagged returns are omitted. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Panel A: 1993 - 2002

	Future annual EPS					Future quarterly EPS			
	1y ahead	2y	3y	4y	5y	1q ahead	2q	3q	4q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\text{Demand}_{i,t}^{3\times3\text{ style}}$	−0.196**	−0.169	−0.124	−0.522	−0.863**	−0.128***	−0.164***	−0.173***	−0.110*
	(0.099)	(0.209)	(0.249)	(0.400)	(0.405)	(0.046)	(0.063)	(0.067)	(0.057)
$\text{Demand}_{i,t}^{6\times6\text{ style}}$	−0.072	−0.013	0.144	−0.217	−0.284	−0.025	−0.047*	−0.062**	−0.061**
	(0.053)	(0.075)	(0.141)	(0.220)	(0.344)	(0.024)	(0.025)	(0.027)	(0.028)
$\text{Demand}_{i,t}^{\text{Idio}}$	0.004	0.011*	0.025**	−0.017	−0.003	−0.003	−0.001	0.004	0.002
	(0.003)	(0.006)	(0.010)	(0.017)	(0.028)	(0.002)	(0.002)	(0.003)	(0.003)
Lagged returns	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month, stock FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	289,508	254,583	86,703	27,726	14,369	278,055	253,929	229,754	200,354
Within R^2	0.016	0.027	0.016	0.023	0.043	0.014	0.019	0.015	0.015

Panel B: 2003 - 2012

	Future annual EPS					Future quarterly EPS			
	1y ahead	2y	3y	4y	5y	1q ahead	2q	3q	4q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\text{Demand}_{i,t}^{3\times3\text{ style}}$	−0.198	−0.077	−0.021	0.382	−0.129	−0.051	−0.131*	−0.079	−0.113
	(0.149)	(0.157)	(0.212)	(0.289)	(0.304)	(0.076)	(0.078)	(0.086)	(0.105)
$\text{Demand}_{i,t}^{6\times6\text{ style}}$	−0.038	−0.011	0.043	−0.278*	−0.186	−0.020	0.026	−0.064	−0.017
	(0.095)	(0.106)	(0.127)	(0.165)	(0.224)	(0.081)	(0.071)	(0.060)	(0.055)
$\text{Demand}_{i,t}^{\text{Idio}}$	0.000	−0.003	−0.014	−0.010	0.007	0.000	0.004	−0.006	−0.005
	(0.009)	(0.007)	(0.013)	(0.017)	(0.018)	(0.004)	(0.005)	(0.005)	(0.007)
Lagged returns	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month, stock FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	277,304	259,121	177,214	78,232	52,919	280,809	273,210	263,469	251,553
Within R^2	0.010	0.018	0.013	0.010	0.012	0.005	0.012	0.012	0.010

Panel C: 2013 - 2022

	Future annual EPS					Future quarterly EPS			
	1y ahead	2y	3y	4y	5y	1q ahead	2q	3q	4q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\text{Demand}_{i,t}^{3\times3\text{ style}}$	−0.048	−0.420	0.154	0.268	0.458	−0.096	−0.291**	−0.062	−0.044
	(0.292)	(0.867)	(0.365)	(0.626)	(0.533)	(0.128)	(0.129)	(0.185)	(0.252)
$\text{Demand}_{i,t}^{6\times6\text{ style}}$	−0.048	0.016	−0.081	0.388	0.249	−0.048	−0.060	−0.025	0.057
	(0.065)	(0.173)	(0.165)	(0.248)	(0.312)	(0.047)	(0.046)	(0.052)	(0.063)
$\text{Demand}_{i,t}^{\text{Idio}}$	0.005	0.018	−0.013	−0.009	−0.020	−0.008	−0.003	0.002	−0.002
	(0.018)	(0.016)	(0.014)	(0.023)	(0.048)	(0.009)	(0.007)	(0.008)	(0.013)
Lagged returns	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month, stock FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	222,360	185,566	134,473	59,305	35,386	237,636	227,648	217,913	205,909
Within R^2	0.007	0.009	0.007	0.006	0.007	0.003	0.010	0.005	0.007

B.2.2 Do the EPS-predicting tests have statistical power?

One may be concerned that our earnings-based regressions lack power. That is, even if demand truly contains information about future earnings, the regressions will still not detect a clear relationship due to misspecification. To examine this concern, we replace demand with the two proxies of CF news in section 4.2. If the regressions are misspecified, we would expect those variables to also not forecast earnings.

We align those two CF news proxies to the monthly forecast time grid in IBES so they can be used in the regressions. We briefly describe them for the readers' convenience. The first variable is standardized unexpected earnings (SUE) from IBES which is defined as quarterly earnings surprises normalized by the standard deviation of analyst forecasts. Because quarterly earnings releases only happen once every three months, the SUE variable is non-zero in one-third of the months and is set to zero in the other two months. The second variable is the firm-level monthly sum of Ravenpack event sentiment score (ESS). Specifically, Ravenpack collects financial news articles on companies and assigns to each article an ESS which indicates how positive or negative the news text is. We sum the ESS of all news released during each IBES monthly period to arrive at our measure. Because there tends to be more news coverage for larger firms, we standardize the ESS measure by firm-year.

Table B.5 shows the regression results of SUE in Panel A and of Ravenpack ESS in Panel B. For ease of comparing coefficients, we standardize both CF news measures to unit variance. Columns (1) and (2) indicate that both news measures have statistically significant predictive power over earnings in the next two years and the predictive coefficient declines by horizon; the results on quarterly EPS in columns (6) to (9) also indicate such a horizon-dependent relationship. In other words, these CF news measures seem to be primarily informative about near-term earnings.

Combined with the fact that lagged stock return controls also have strongly significant coefficients (Table 2), we conclude that our test does not lack power: variables that truly predict earnings will end up having a positive coefficient. Therefore, the lack of a clear relationship between our demand measure and future EPS indicates that demand does not contain information about future EPS.

Table B.5. Predicting earnings using cash flow news proxies

This table reports forecasting regressions similar to that in Table 2 but using alternative independent variables. Panel A uses standardized unexpected earnings (SUE) from IBES and Panel B uses firm-level event sentiment score (ESS) from Ravenpack. To make regression coefficients easy to compare, both SUE and ESS are standardized to have unit variance and the dependent variable is in percent. To save space, the regression coefficients on lagged stock returns are not reported. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Panel A: standardized unexpected earnings

	Future annual EPS					Future quarterly EPS			
	1y ahead	2y	3y	4y	5y	1q ahead	2q	3q	4q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\text{SUE}_{i,t}$	0.125***	0.068***	0.015	0.018	−0.031	0.051***	0.044***	0.043***	0.037***
	0.010	0.011	0.015	0.023	0.034	0.006	0.004	0.005	0.005
Lagged returns	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month, stock FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	741,521	662,782	388,811	162,208	101,265	765,610	730,787	691,450	641,989
Within R^2	0.014	0.019	0.011	0.010	0.012	0.008	0.014	0.011	0.011

Panel B: event sentiment score

	Future annual EPS					Future quarterly EPS			
	1y ahead	2y	3y	4y	5y	1q ahead	2q	3q	4q
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\text{ESS}_{i,t}$	0.167***	0.090***	0.025	−0.060*	−0.091*	0.081***	0.082***	0.080***	0.055***
	0.015	0.022	0.027	0.032	0.048	0.009	0.008	0.009	0.010
Lagged returns	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month, stock FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	729,021	635,121	375,169	156,159	96,351	753,324	724,476	689,785	642,460
Within R^2	0.010	0.013	0.008	0.006	0.008	0.004	0.007	0.006	0.005

B.3 Sizing price multipliers justified by information

Section 2.4 estimates the predictive power that demand has on future EPS. This section uses those predictive coefficients to estimate the price multiplier that can be justified by the information in demand.

We start with a present value formula for the per-share fundamental value of stock i:

$$V_i = \sum_{h=1}^{\infty} \frac{\tau \cdot X_i^h}{(1+r)^h} \tag{A4}$$

where τ is the earnings payout ratio, r is the appropriate risky discount rate, and X_i^h is the h-year ahead EPS. Because discount rates and payout ratios are slow-moving variables, over a short period such as a week, stock price variation changes should be primarily driven by earnings beliefs:

$$P_{i,t} = \sum_{h=1}^{\infty} \frac{\tau \cdot \tilde{E}_t(X_i^h)}{(1+r)^h}$$

where $\tilde{E}_t(\cdot)$ denotes market expectations. Suppose that demand $\{\text{Demand}_{i,t}^l\}_{\text{level } l l}$ is realized, and let $\tilde{E}_{t+}(\cdot)$ denote the updated market expectation after incorporating the EPS-relevant information in demand. This causes return movement of

$$\frac{P_{i,t+} - P_{i,t}}{P_{i,t}} = \sum_{h=1}^{\infty} \frac{\tau}{(1+r)^h} \cdot \frac{\tilde{E}_{t+}(X_i^h) - \tilde{E}_t(X_i^h)}{P_{i,t}}. \tag{A5}$$

Equation (A5) has a natural connection with the regressions in Table 2 which, using the notation here, estimate b_h^l for each horizon $h = 1, ..., 5$:

$$\frac{X_i^h - \tilde{E}_t(X_i^h)}{P_{i,t}} = a + \sum_{\text{level } l} b_h^l \cdot \text{Demand}_{i,t}^l + \epsilon_{i,t}.$$

Under the assumption of rational expectations — that is, beliefs are updated according to *genuine* information — the information-justified price multipliers for each level l are given by:

$$M_l = \sum_{h=1}^{\infty} \frac{\tau \cdot b_h^l}{(1+r)^h} \tag{A6}$$

To be lenient, we assume a discount rate of $r = 0$ and a payout ratio of $\tau = 1$, so equation (A6) simplifies to $\sum_{h=1}^{\infty} b_h^l$ which amounts to adding up forecasting coefficients.

If we take the point estimates for $h = 1$ to 5 in Table 2, this leads to point estimates for M_l of -0.229, -0.036, and -0.019 for the 3×3 style, 6×6 style, and idiosyncratic levels, respectively. Assuming independence across estimated forecasting coefficients, the standard errors are 0.444, 0.264, and 0.022, respectively. Therefore, at 1% confidence (t-stat = 2.576), we obtain upper bounds of information-justified price multipliers of 0.915, 0.644, and 0.036, respectively, at the three levels we are interested in. These upper bounds are significantly lower than the empirical estimates of price multipliers in Table 3. Therefore, as far as the next five years of earnings are concerned, CF information cannot justify the empirically observed price effects of demand.

B.4 Omitted variable at the market-level: S&P futures order flow

As discussed in section 2.2, previous studies have found that S&P futures order flow can account for a large fraction of market-level price movements (Deuskar and Johnson, 2011). Thus, directly using our demand measure to study market-level price impact suffers from an omitted variable bias. Further, if our demand measure is positively correlated with the unobserved S&P futures flow, we will overestimate the market-level price multiplier. Concretely, suppose the true market-level relationship is:

$$r_t^{\text{market}} = a + b \cdot \text{Demand}_t^{\text{market}} + c \cdot \text{FuturesFlow}_t + \epsilon_t$$

where r_t^{market} is the market return, $\text{Demand}_t^{\text{market}}$ is the value-weighted average of our demand measure, and FuturesFlow_t is the S&P futures flow. If we omit futures flows, our estimated price multiplier will be biased by the following amount:

$$\begin{aligned}\hat{b} &= \frac{Cov(r_t, \text{Demand}_t^{\text{market}})}{Var(\text{Demand}_t^{\text{market}})} \\ &= b + \underbrace{c \cdot \frac{Cov(\text{Demand}_t^{\text{market}}, \text{FuturesFlow}_t)}{Var(\text{Demand}_t^{\text{market}})}}_{\text{bias}}\end{aligned} \tag{A7}$$

The authors of Deuskar and Johnson (2011) generously agreed to help quantify this bias.

While they cannot share their data, they agreed to run a regression of their S&P future flow (FuturesFlow_t) on $\text{Demand}_t^{\text{market}}$. The full sample result is reported in column (1) of Table B.6. As suspected, our demand is indeed positively correlated with futures order flow, with 1% higher demand associated with approximately $1\% \times 116,122,000 \approx 1,161,220$ more future contracts bought. The relationship is statistically significant at the 1% level. Columns (2) through (4) estimate the same regression using subsamples. While the results have larger standard errors and lose statistical significance in the first two years, the relationship is positive in each subsample and the coefficient does not vary too much over time.

Table B.6. Projecting S&P futures order flow onto our demand measure

With help from the authors of Deuskar and Johnson (2011), we regress daily S&P futures order flow on our value-weighted market-level demand measure over their sample period of February 13, 2006 to January 9, 2019. Column (1) estimates the regression using the full sample while columns (2) through (4) use sub-samples by year. The year 2009 is omitted due to insufficient data. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Independent variable: S&P futures order flow (thousand contracts)				
	Full sample	Sub-samples 2006	2007	2008
	(1)	(2)	(3)	(4)
$\text{Demand}_t^{\text{market}}$	$116,122^{***}$	$84,253$	$78,248$	$140,402^{**}$
	(34,727)	(55,220)	(62,237)	(59,995)
Obs	731	223	250	252
Adj R^2	0.014	0.006	0.002	0.018

We can now use equation (A7) to conduct a back-of-the-envelope quantification of the bias if we directly estimate the market-level price multiplier using our demand measure. Table 2 in Deuskar and Johnson (2011) estimates the price impact of futures flow (variable c in equation (A7)) to be approximately 2.4 basis points per thousand futures contracts.[8] Multiplying this by the full sample estimate in Table B.6 indicates that the market price multiplier estimate will be biased by $(2.4 \times 10^{-5}) \times 116,112 = 27.9$, and taking into account estimation errors in Table B.6 results in a wide 95% confidence interval of [11.5, 44.2]. We

[8]More specifically, the first row in their Table 2 estimates the average price impact to be 0.32 index points per 1,000 contracts. The average value of S&P 500 index is 1334 points during their sample period, so this translates to $c = 0.32/1334 \approx 2.4$ basis points per 1,000 contracts.

cannot take the exact estimate literally as it is only based on a short sample, but this does indicate that the omitted variable problem is severe, and the market-level price multiplier can be severely upward biased.

Using our demand measure to estimate market-level multipliers. We now confirm that we indeed find overly large market-level price multipliers when using our demand measure. In Table B.7, we report time-series regressions of weekly value-weighted market returns on our value-weighted demand measure. Column (1) uses the full sample and finds a market-level price multiplier estimate of 25.73. Columns (2) through (5) report results by equal-length subperiods. While the estimate fluctuates somewhat over time, all estimates are above 20.

Table B.7. Estimating market-level multipliers with our demand measure

We regress weekly value-weighted market return onto our value-weighted demand measure. Column (1) reports regression results for the full sample. Columns (2) through (4) report estimates using equal-length sub-periods. In columns (6) and (7), we estimate the regression on the Deuskar and Johnson (2011) sample period and its complement, respectively. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable: weekly market return						
	Full sample	Equal-length sub-samples 1993-1999	2000-2007	2009-2015	2016-2022	DJ (2011) period	Non-DJ (2011) period
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\text{Demand}_t^{\text{market}}$	25.73***	24.97***	21.07***	28.38***	29.89***	32.85***	24.45***
	(0.99)	(1.30)	(1.91)	(1.65)	(4.05)	(3.16)	(1.04)
Obs	1,511	351	405	401	354	148	1,363
Adj R^2	0.310	0.512	0.230	0.425	0.132	0.421	0.290

These estimates appear too high when compared to existing studies. Using a granular instrument approach, Gabaix and Koijen (2022) estimates the market-level price multipliers to be around 5; they also cite a few studies that find estimates in a similar range. In column (6) of Table B.7, we find that the estimated multiplier is 32.85 during the sample period of Deuskar and Johnson (2011) which is around the same order of magnitude as our estimated bias of 27.9 based on Table B.6. Overall, the results are consistent with the idea of a severe omitted variable bias, and thus we conclude that our measure should not be used to estimate market-level price multipliers.

C Additional results on price multipliers

Section 3 shows that price multipliers tend to be larger at more aggregate levels. This section presents additional empirical estimates.

C.1 Price effects are persistent

While we estimate the main result in Table 3 at a weekly frequency, the result would be of less interest for asset pricing purposes if it reverts soon after a week. Short-lived price impacts would not lead to meaningful explanatory power over asset prices.

To study the persistence of price impacts, we modify regression (9) to use longer-horizon returns. Specifically, for each horizon $h = 0, ..., 8$ weeks, we estimate:

$$\begin{aligned}\mathrm{Ret}_{i,t+t+h} = a + M_h^{3\times3} \cdot \mathrm{Demand}_{l1(i),t}^{3\times3} + M_h^{6\times6} \cdot \mathrm{Demand}_{l2(i),t}^{6\times6} \\ + M_h^{\mathrm{Idio}} \cdot \mathrm{Demand}_{i,t}^{\mathrm{Idio}} + \mathrm{Controls}_{i,t} + \epsilon_{i,t}\end{aligned} \tag{A8}$$

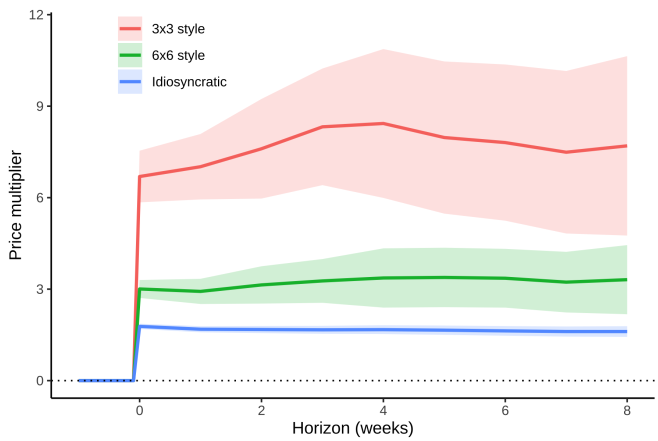
where, for each level $l \in \{3\times3, 6\times6, \mathrm{Idio}\}$, coefficients $\{M_h^l\}_{h=0}^8$ traces out the impulse response of demand-based price effects. Because we are now estimating regressions with overlapping time periods, we estimate standard errors using the Driscoll and Kraay (1998) procedure which captures both cross-sectional and time-series correlations, and the number of lags in standard error calculation is equal to the number of weeks spanned by the dependent variable.

Panel (a) of Figure C.5 plots the estimated price multipliers $\{M_h^l\}_{h=0}^8$. For visualization purposes, the lines start at $M_{-1}^l = 0$ so the plots can be interpreted as impulse responses. We find no evidence of price impacts reversing. Panels (b) and (c) plot the price effects estimated on the first and second halves of the sample, respectively. While price multipliers are generally smaller in the later period, we do not find evidence of price impact reversals in either period.

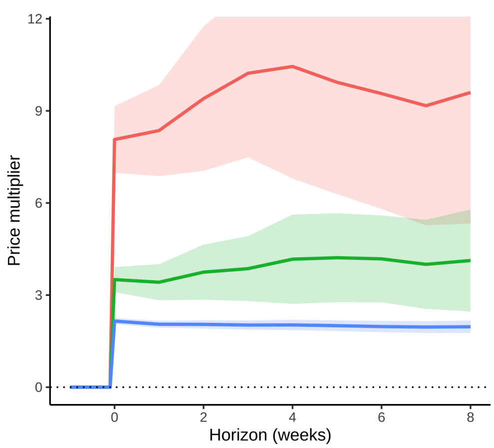
Figure C.5. Long-run price effects of demand

This figure plots the impulse response of demand-induced price effects with 95% confidence bands. We estimate these price multipliers using panel regressions of cumulative stock return over weeks t to $t+h$ on decomposed week-t demand for $h = 0, ..., 8$. Standard errors are computed using the Driscoll and Kraay (1998) procedure. Panel (a) plots estimates for the full sample. Panels (b) and (c) plot results estimated for the first and second half of the sample, respectively.

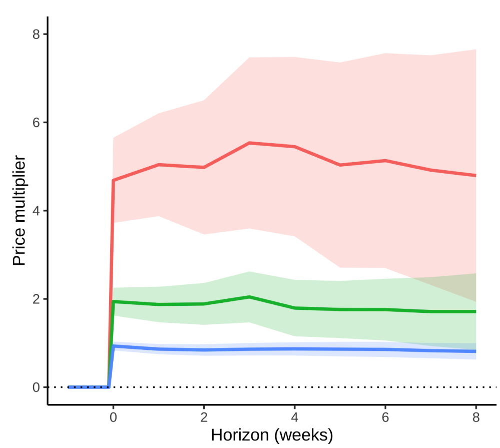
(a) Full sample



(b) Sub-sample: 1993 - 2007



(c) Sub-sample: 2008 - 2022



C.2 Robustness of the main results

This section presents several robustness checks of the findings in section 3.1. We are particularly interested in whether the findings are robust to alternative demand decomposition methods.

C.2.1 Demand decomposition using other stock characteristics

To test whether the results reported in Table 3 are specific to size- and BM-based decomposition, we compute alternative demand decompositions based on pairwise combinations of size, BM, momentum, investment, and operating profitability. This yields ten possible combinations. We then estimate price multipliers using weekly regressions and report the results in Table C.8. For all ten specifications, we find the same conclusion that multipliers are larger for more aggregate demand components. In the bottom panel of the table we show that, for all specifications, the price multiplier for the 3×3 style level is larger than the multiplier at the 6×6 style level which, in turn, is larger than the multiplier for the idiosyncratic level demand. All differences are statistically significant at the 1% level.

C.2.2 Demand decomposition using flexible combinations of characteristics

Even though Table C.8 confirms that the main results are robust to using alternative stock characteristics, one may still worry that our results are sensitive to our double-sort decomposition procedure or to our choice of characteristics. It would be comforting to know that our results are robust to demand decomposing using many characteristics in a bottom-up fashion without manual selection. For this purpose, we adopt the instrumented principal component analysis (IPCA) approach which builds a factor model of the cross-section of stocks where factors are constructed in a bottom-up fashion by combining many stock characteristics (Kelly et al., 2019; Kelly, Pruitt, and Su, 2020).

The IPCA method. We briefly summarize the IPCA methodology for the reader's convenience. We follow Kelly et al. (2019) to estimate a dynamic factor model for returns:

$$r_{i,t+1} = z_{i,t}^{'} \Gamma_{\beta} f_{t+1} + \epsilon_{i,t+1}$$

Table C.8. Using other characteristics for demand decomposition

This table presents a robustness check of the main results in Table 3. We use pairwise combinations of size (me), book-to-market (bm), momentum (mom), operating profitability (prof), and investment (agr) stock characteristics to decompose demand into three components via the procedure described in section 2.3. We then use panel regressions to estimate the price multipliers associated with each of the three components. In all regressions we control for week fixed effects and the list of controls included in Table 3. Panel A reports regression results and Panel B reports pairwise differences in multiplier estimates. The standard errors are clustered by week and stock and reported in parentheses. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Panel A: Price multiplier estimates										
	Weekly stock returns ($\text{Ret}_{i,t}$)									
Characteristics	me-bm	me-agr	me-prof	me-mom	bm-agr	bm-prof	bm-mom	agr-prof	agr-mom	prof-mom
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\text{Demand}^{3\times3}_{i,t}$	6.976***	5.543***	5.012***	7.483***	5.280***	5.008***	7.152***	4.098***	6.795***	6.035***
	(0.450)	(0.305)	(0.297)	(0.430)	(0.361)	(0.400)	(0.454)	(0.236)	(0.419)	(0.359)
$\text{Demand}^{6\times6}_{i,t}$	3.121***	2.708***	2.494***	3.786***	2.265***	2.193***	2.715***	2.046***	2.923***	2.625***
	(0.153)	(0.119)	(0.109)	(0.174)	(0.092)	(0.081)	(0.115)	(0.085)	(0.131)	(0.102)
$\text{Demand}^{\text{Idio}}_{i,t}$	1.732***	1.731***	1.707***	1.725***	1.743***	1.714***	1.748***	1.706***	1.738***	1.711***
	(0.036)	(0.038)	(0.038)	(0.036)	(0.038)	(0.038)	(0.037)	(0.040)	(0.038)	(0.038)
Week FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Other controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs	3,922,489	3,703,825	2,939,004	4,057,402	3,682,660	2,912,079	3,889,511	2,760,987	3,696,353	2,914,086
Within R^2	0.038	0.037	0.036	0.039	0.036	0.036	0.038	0.035	0.037	0.037
Panel B: Price multiplier differences										
$\text{Demand}^{3\times3}_{i,t}$	3.856***	2.835***	2.518***	3.697***	3.015***	2.815***	4.437***	2.052***	3.872***	3.410***
$-\text{Demand}^{6\times6}_{i,t}$	(0.475)	(0.327)	(0.316)	(0.464)	(0.373)	(0.408)	(0.468)	(0.251)	(0.439)	(0.373)
$\text{Demand}^{6\times6}_{i,t}$	1.388***	0.978***	0.787***	2.062***	0.522***	0.479***	0.967***	0.340***	1.185***	0.914***
$-\text{Demand}^{\text{Idio}}_{i,t}$	(0.157)	(0.125)	(0.116)	(0.178)	(0.099)	(0.089)	(0.120)	(0.094)	(0.137)	(0.109)

where $r_{i,t+1}$ is the monthly excess return on stock i in month $t+1$, $z_{i,t}$ is an L-vector of lagged stock characteristics, Γ_β is a $L \times K$ matrix of coefficients, and f_{t+1} is an K-vector of factors. In this exercise, K is set to L because we want to estimate a comprehensive list of factors that capture systematic risk variations at all levels. Since there are 5,753 stocks, overfitting is not a concern because we are capturing most return variations with a much smaller number of factors than the number of stocks. The coefficients Γ_β and the hidden factors f_{t+1} are estimated by minimizing the average squared distance between the explained returns, $z'_{i,t}\Gamma_\beta f_{t+1}$, and actual return realizations:

$$(\hat{\Gamma}_\beta, \hat{f}_{t+1}) = \arg\min_{\Gamma_\beta, f_{t+1}} \sum_{t=1}^{T-1} \sum_{i=1}^{N} (r_{i,t+1} - z'_{i,t}\Gamma_\beta f_{t+1})^2.$$

Kelly et al. (2019) show that the resulting factors can be interpreted as managed portfolios whose weights are transformations of the stock characteristics. To see this, examine the first-order conditions for the factors:

$$\hat{f}_{t+1} = \left(\hat{\Gamma}'_{\beta} Z'_t Z_t \hat{\Gamma}_{\beta}\right)^{-1} \hat{\Gamma}'_{\beta} Z'_t r_{t+1}$$

where $r_t = (r_{1,t}, ..., r_{N,t})'$ and $Z_t = (z_{1,t}, ..., z_{N,t})'$.

Demand decomposition To replicate the Kelly et al. (2019) implementation as closely as possible, we simply took their stock return and characteristics data and their code, both of which are made available on Seth Pruitt's website. We use the overlapping part of their data with ours, and the merged dataset runs from 1993 through 2014. Their data include the 36 characteristics from Freyberger, Neuhierl, and Weber (2020) normalized to uniform distributions over [-0.5, 0.5], as described in Kelly et al. (2019). We use their algorithm to estimate $K = 36$ factors, which gives us the same number of dimensions as the number of characteristics, and then rank the factors in decreasing order of return-explanatory power.

We then use the estimated IPCA model to decompose weekly demand. Specifically, after estimating $\hat{\Gamma}_{\beta}$, we regress demand on the characteristics-instrumented loadings to obtain the corresponding demand factors: $f_{d,t+1} = (\hat{\Gamma}'_{\beta} Z'_t Z_t \hat{\Gamma}_{\beta})^{-1} \hat{\Gamma}'_{\beta} Z'_t d_{t+1}$, where d_{t+1} is the vector of demand for all stocks in week $t+1$. For the demand factors, $Z_{i,t} \hat{\Gamma}_{\beta} f_{d,t+1}$, we split the 36 factors into two equal-size groups: the "more aggregate" group is the sum of demand factors 1 to 18 and the "less aggregate" group includes factors 19 to 36. We define the idiosyncratic demand as the residual after subtracting all factor demand components from the raw demand.

Price multipliers Having decomposed demand into three components with varying levels of aggregation, we then estimate their price multipliers using the same type of panel regression as in equation (9):

$$\text{Ret}_{i,t} = M^{\text{More aggregate}} \cdot \text{Demand}_{i,t}^{\text{More aggregate}} + M^{\text{Less aggregate}} \cdot \text{Demand}_{i,t}^{\text{Less aggregate}}$$
$$+ M^{\text{Idiosyncratic}} \cdot \text{Demand}_{i,t}^{\text{Idiosyncratic}} + \text{Controls}_{i,t} + \epsilon_{i,t} \tag{A9}$$

where the controls and the standard error clustering match those in equation (9). The full-sample regression is reported in column (1) of Table C.9. As is the case with the results based on bivariate characteristics–based demand decomposition, the more aggregate demand components are associated with larger price multipliers. Columns (2) through (4) report sub-sample estimates and the multiplier orders are preserved. In columns (5) to (8), we also estimate market cap-weighted regressions and the results are qualitatively similar. Overall, we conclude that our main result is robust to using more stock characteristics to decompose demand with a bottom-up approach.

C.3 Results based on Flow-Induced Trading

In this section, we use the flow-induced trading (FIT) instrument in Lou (2012) to study how price multipliers vary across levels of aggregation. A long literature has argued that mutual fund flows represent uninformed trading (e.g. Coval and Stafford, 2007; Frazzini and Lamont, 2008; Ben-David, Li, Rossi, and Song, 2021b). Further, in response to fund flows, mutual fund managers mechanically adjust their positions, and these flows appear to create price pressures at both the stock- and the style levels (e.g. Teo and Woo, 2004; Lou, 2012; Li, 2022; Huang et al., 2022; Ben-David et al., 2022). Therefore, we conduct a similar exercise to examine whether FIT also commands larger price multipliers at more aggregate levels.

We obtain monthly mutual fund flows from CRSP and merge them with lagged quarterly holdings data from Thomson Reuters S12. We follow Lou (2012) to compute FIT for each

Table C.9. Price multipliers: IPCA-decomposed demand components

This table presents a robustness check of Table 3. As described in section C.2.2, we use the instrumented principal component analysis (IPCA) method in Kelly et al. (2019) to decompose demand into three components. The IPCA method builds a dynamic factor model of stocks where each factor can be interpreted as a managed portfolio whose weights are transformations of stock characteristics. We estimate a model with 36 factors following their method, rank them in decreasing order of return explanatory power, and use it to decompose weekly demand. The first half of the factors are summed into the "more aggregate" component; the second half of the factors are summed into the "less aggregate" component; the residual is termed the idiosyncratic component. We then estimate panel regressions of weekly stock returns on these IPCA-decomposed demand components with the same set of controls as with Table 3. Standard errors, clustered by week and stock, are reported in parentheses. In the first four columns we report equal-weighted regressions and in the last four columns we report market cap-weighted regressions. In columns (1) and (5) we report full-sample estimates while in the other columns we report subsample estimates. Because we merge the data in Kelly et al. (2019) with ours, the data period ends in 2014 and the data contains fewer stocks. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable: Weekly stock returns ($\text{Ret}_{i,t}$)							
	Equal weighted				Value weighted			
	Full sample	1993-2000	2001-2007	2008-2014	Full sample	1993-2000	2001-2007	2008-2014
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Demand}_{i,t}^{\text{More aggregate}}$	5.501***	9.073***	4.379***	6.102***	6.082***	9.974***	4.977***	6.196***
	(0.270)	(0.573)	(0.274)	(0.489)	(0.308)	(0.593)	(0.372)	(0.582)
$\text{Demand}_{i,t}^{\text{Less aggergate}}$	4.222***	8.235***	3.340***	4.285***	4.797***	8.795***	3.797***	5.155***
	(0.204)	(0.435)	(0.233)	(0.416)	(0.249)	(0.477)	(0.311)	(0.566)
$\text{Demand}_{i,t}^{\text{Idiosyncratic}}$	1.821***	2.886***	1.364***	0.679***	2.255***	4.379***	1.695***	0.153**
	(0.046)	(0.052)	(0.055)	(0.058)	(0.112)	(0.165)	(0.122)	(0.074)
Week FE	Y	Y	Y	Y	Y	Y	Y	Y
Stock FE	Y	Y	Y	Y	Y	Y	Y	Y
Other controls	Y	Y	Y	Y	Y	Y	Y	Y
Obs	1,889,049	691,289	664,949	532,811	1,889,049	691,289	664,949	532,811
Adj R^2	0.239	0.211	0.211	0.360	0.285	0.287	0.255	0.388

stock i in month t:

$$\text{FIT}_{i,t} = \frac{\sum_{\text{fund } j} \text{SharesHeld}_{i,j,t-1} \cdot \text{Flow}_{j,t}}{\text{SharesOutstanding}_{i,t-1}}$$

where $\text{SharesHeld}_{i,j,t-1}$ is the lagged number of stock i shares held by fund j and monthly fund flow is computed conventionally: $\text{Flow}_{j,t} = \frac{\text{AUM}_{j,t} - \text{AUM}_{j,t-1} \cdot \text{FundReturn}_{j,t}}{\text{AUM}_{j,t-1}}$. There are two main differences from Lou (2012). First, the denominator is lagged shares outstanding, rather than the number of shares held by mutual funds, because we want to estimate price multipliers. Second, for simplicity, we assume both in- and out-flows are passed on one-to-one to fund trades, rather than use differential scaling factors. Li (2022) finds that this specification difference makes little difference.

We then follow the methodology in section 2.3 to decompose FIT into three components,

$$\text{FIT}_{i,t} = \text{FIT}^{3\times3}_{l1(i),t} + \text{FIT}^{6\times6}_{l1(i),t} + \text{FIT}^{\text{Idio}}_{i,t}.$$

We then estimate price multipliers using monthly panel regressions of stock returns on decomposed FIT components with time fixed effects and the same set of controls as in Table 3. The resulting prima facie multiplier estimates ($\tilde{M}$) are presented in columns (1) of Table C.10. Similar to the results based on our demand measure, estimated multipliers are larger at more aggregate levels.

Table C.10. Price multipliers estimates using flow-induced trading

We use monthly panel regressions to estimate price multipliers using fund flow-induced trading (FIT) as defined in Lou (2012). Column (1) presents the "prima facie" multiplier estimates ($\tilde{M}$); the regression specification is the same as in Table 3 at monthly frequency. Column (3) presents estimates of flow scaling factors estimated at each level where we regress $\text{FIT}^{\text{lev}}_{i,t\to t+17}$ on $\text{FIT}^{\text{lev}}_{i,t}$ and report the coefficients β. Column (2) reports the scaling factor-adjusted multiplier $M = \tilde{M}/\beta$. The last two columns present pairwise differences of multiplier estimates for both the prima facie and adjusted multiplier estimates. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

	Multiplier estimates				Differences	
	Prima facie ($\tilde{M}$)	Adjusted ($M = \tilde{M}/\beta$)	FIT scaling factor (β)		Prima facie ($\tilde{M}$)	Adjusted ($M = \tilde{M}/\beta$)
	(1)	(2)	(3)		(1d)	(2d)
$\text{FIT}^{3\times3}_{i,t}$	10.852***	2.854***	3.802***	$\text{FIT}^{3\times3}_{i,t} - \text{FIT}^{6\times6}_{i,t}$	4.193	1.139*
	(1.934)	(0.509)	(0.608)		(2.655)	(0.692)
$\text{FIT}^{6\times6}_{i,t}$	6.659***	1.715***	3.882***	$\text{FIT}^{6\times6}_{i,t} - \text{FIT}^{\text{Idio}}_{i,t}$	2.827	0.685
	(1.820)	(0.469)	(0.351)		(1.843)	(0.475)
$\text{FIT}^{\text{Idio}}_{i,t}$	3.832***	1.030***	3.720***	$\text{FIT}^{3\times3}_{i,t} - \text{FIT}^{\text{Idio}}_{i,t}$	7.020***	1.824***
	(0.291)	(0.078)	(0.106)		(1.955)	(0.515)
Month FE	Y					
Controls	Y					
Obs	887,949					
Within R^2	0.009					

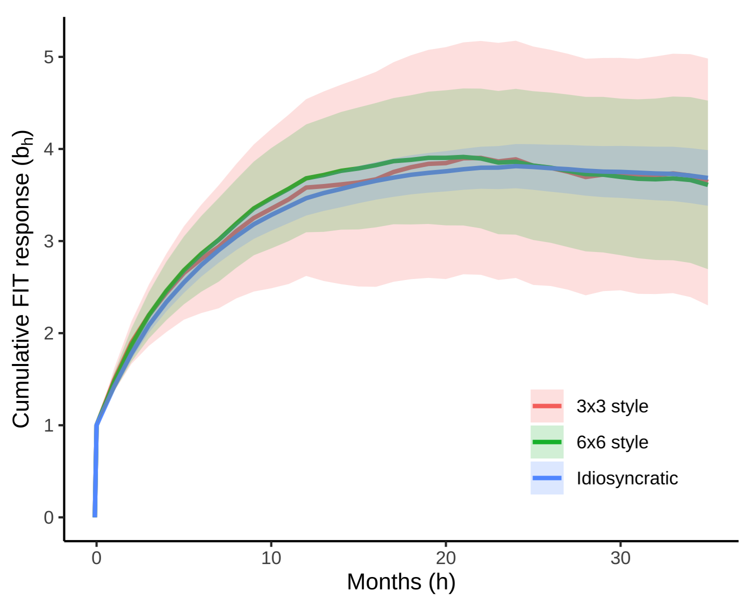
The magnitudes of FIT-based multiplier estimates are larger than estimates based on our demand measure. However, this is largely due to FIT being more persistent, a point that is discussed in more detail in Appendix F.2. Unlike our demand measure which exhibits little autocorrelation at weekly or slower frequencies, FIT is highly autocorrelated because both mutual fund holdings and flows (e.g. Coval and Stafford, 2007) are persistent. To gauge the degree of autocorrelation, at each level of aggregation, we estimate for horizons of $h = 1, ..., 36$ months:

$$\text{FIT}_{i,t\to t+h-1} = b_h^{\text{lev}} \cdot \text{FIT}_{i,t} + \sum_{t=1}^{T} \eta_t^{\text{lev}} + \epsilon_{i,t\to t+h-1}^{\text{lev}}$$

where $\{\eta_t^{\text{lev}}\}_{t=1}^{T}$ are month fixed effects and we cluster standard errors by month. Figure C.6 plots the impulse responses traced out by the coefficients $\{b_h^{\text{lev}}\}_{h=1}^{36}$. We see that all components of FIT are persistent for around 18 months. We these choose to use the 18-month coefficient as our estimate for β and report them in column (3) of Table C.10. For all levels of aggregation, each 1 unit of FIT implies another 2.7 to 2.9 units more to come in the subsequent months.

Figure C.6. Decomposing demand into components using stock characteristics

We estimate the persistence of different components of FIT by regressing $\text{FIT}_{i,t\to t+h-1}$ on $\text{FIT}_{i,t}$ with time fixed effects. As we vary $h = 1, ..., 36$, the resulting regression coefficients trace out the impulse responses which are plotted in the figure. The error bands represent 95% bootstrapped confidence intervals.



The adjusted multipliers, $M = \tilde{M}/\beta$, are presented in column (2) of Table **??**. Both before and after adjustment, the multipliers appear larger at more aggregate levels, albeit with much larger standard errors than estimates based on our demand measure. Overall, FIT has weaker power in explaining returns as the full sample within-period R^2 is around one-quarter that of our demand measure in Table 3. Columns (1d) and (2d) report the pairwise differences in multipliers. With larger standard errors, the differences between adjacent demand components are no longer statistically significant, but the difference between the 3×3 and idiosyncratic levels are still statistically significant at the 1% level. Overall, the results based on FIT are consistent with the idea that multipliers form a continuum.

D The information-based alternative hypothesis

This appendix section supplements analyses in sections 4.2 and 4.3. Appendix D.1 explores the two directional news measures in section 4.2. Appendix D.2 explores the non-directional news indicators in section 4.3. Appendix D.3 explains why sorting on contemporaneous realized volatility leads to spurious results.

D.1 Stock-level directional news measures

This section provides additional details on section 4.2. We first describe how the two measures are computed and then study their properties.

News measure construction. The first measure is standardized unexpected earnings (SUE) from Ravnepack which captures the earnings surprise in quarterly releases. Specifically, SUE is defined as the difference between realized EPS and consensus analyst forecast prior the the release, and the difference is normalized by the standard deviation of analyst forecasts. Previous studies have shown that SUE explains sizeable stock return variation around earnings periods (e.g. Livnat and Mendenhall, 2006; Hirshleifer, Lim, and Teoh, 2009).

To capture more news during weeks without earnings releases, our second variable summarizes media news sentiment using data from Ravenpack. Specifically, Ravenpack collects financial news articles from a variety of sources, links them to the relevant stocks, and assigns a relevance score $\in [0, 100]$ to indicate how relevant the article is to the stock. We focus attention on articles with relevance score ≥ 75 to avoid less relevant news coverage. Further, Ravenpack assigns an event sentiment score (ESS $\in [-1, 1]$) to each article based on how positive or negative is the textual tone. We sum up the ESS of all articles by stock-week to arrive at our measure. As a validation of the usefulness of this measure, Boudoukh et al. (2019) find that Ravenpack news measures can explain stock returns.

One issue with the Ravenpack ESS measure is that news coverage is more extensive for large firms and for the later part of the sample. This requires us to adjust the scale. Therefore, we standardize the ESS measure by firm-year to zero mean and unit variance. For comparability, we also standardize the SUE measure similarly.

To visualize the standardized SUE and ESS measures, Figure D.7 plots them for the company AT&T in the year 2021. Panel (a) plots SUE which is non-zero only on weeks with earnings releases that are signified by blue markers. The plot shows that AT&T had a negative earning surprise on the first release and larger positive surprises in the three subsequent releases. Panel (b) plots the ESS measure which varies continuously, indicating that there can be substantial value-relevant news releases on non-earnings weeks.

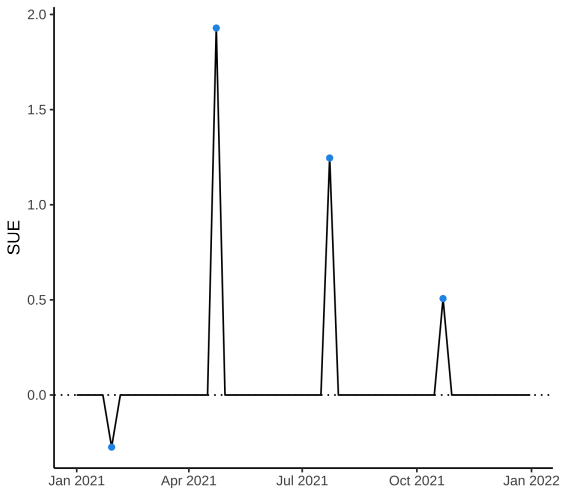
News measures explain stock returns. Consistent with prior literature, we confirm that both news measures have significant explanatory power over stock returns. In Panel (a) of Figure D.8, we sort the sample of earnings weeks into deciles by SUE and plot the average stock return over market. We find a monotonic relationship between SUE and realized stock returns with the top-to-bottom SUE decile difference explaining a spread of 9.6% in weekly stock returns. In Panel (b), we sort the full sample into ESS deciles and find that the top-to-bottom decile difference explains a spread of 4.4% in weekly return.

(Lack of) relationship with demand. We merge the standardized SUE and ESS news measures with our weekly demand data. We use the same methodology as in section 2.3 to

Figure D.7. SUE and stock-level ESS: sample data

We use standardized unexpected earnings (SUE) and Ravenpack event sentiment score (ESS) to measure weekly firm-level news. Both measures are standardized by year-stock. To illustrate the data, Panels (a) and (b) plot the SUE and ESS of AT&T in 2021, respectively. The blue dots mark the four weeks with quarterly earnings releases. The horizontal dashed line marks zero.

(a) SUE



(b) ESS

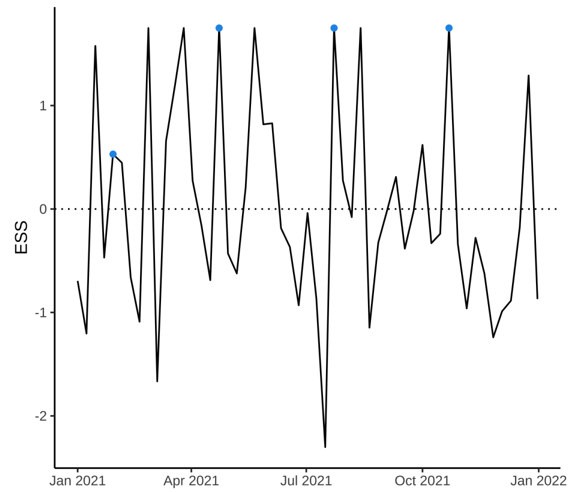
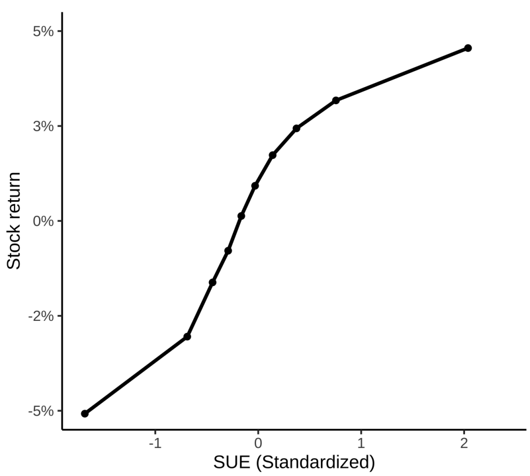


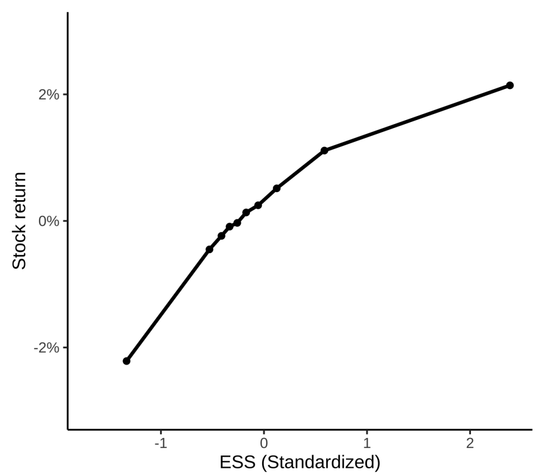
Figure D.8. Firm-level news measures and stock returns

This figure examines the relationship between stock returns and two stock-level news measures. Panel (a) examines standardized unexpected earnings (SUE) and panel (b) examines Ravenpack event sentiment score (ESS). Both measures are standardized by firm-year. In both panels, we sort the sample by the examined news measure into deciles and plot the average weekly stock return in excess of the market.

(a) Standardized Unexpected Earnings (SUE)



(b) Event Sentiment Score (ESS)



decompose SUE and ESS into three components with declining levels of aggregation. Panel A of Table D.11 presents summary statistics for the merged sample. Panel B presents the average cross-sectional correlation between the demand and news measures. The results indicate that demand has little correlation with SUE or ESS at any level, which is consistent with the finding that controlling for SUE and ESS in Table 5 does not affect the demand-based price multipliers.

Table D.11. Summary statistics for firm-level news measures

In this table we report summary statistics for the weekly sample that has been merged with IBES standardized earnings surprise (SUE) and Ravenpack event sentiment score (ESS). SUE and ESS are both standardized and decomposed into three components in the same way we decompose demand (see section 2.3). Panel A reports summary statistics. In column (1) we report the average number of stocks per week. In columns (2) and (3) we report the means and standard deviations, and in the last three columns we report the 5%, 50%, and 95% percentile distributions of each variable. Panel B reports the average cross-sectional correlation between demand components and decomposed components of SUE and ESS. Specifically, for any two variables, we compute their correlation by week and then report the average over all weeks.

Panel A: summary statistics

		Obs	Mean	StDev	Percentiles		
					5%	50%	95%
		(1)	(2)	(3)	(4)	(5)	(6)
Demand (%)	3×3 style	1,943	0.00	0.08	-0.12	0.00	0.11
	6×6 style	1,943	0.00	0.07	-0.11	0.00	0.11
	Idiosyncratic	1,943	0.00	0.58	-0.79	0.00	0.81
SUE (%)	3×3 style	1,943	-0.02	0.11	-0.22	0.00	0.13
	6×6 style	1,943	0.00	0.13	-0.20	0.00	0.20
	Idiosyncratic	1,943	0.00	1.02	-0.45	0.00	0.31
ESS (%)	3×3 style	1,943	0.00	0.16	-0.29	0.01	0.23
	6×6 style	1,943	0.00	0.14	-0.21	0.00	0.21
	Idiosyncratic	1,943	0.00	0.96	-0.89	-0.13	1.67

Panel B: correlation matrix

	$\text{SUE}^{3\times3}$	$\text{SUE}^{6\times6}$	SUE^{Idio}	$\text{ESS}^{3\times3}$	$\text{ESS}^{6\times6}$	ESS^{Idio}
$\text{Demand}^{3\times3}$	-0.022	0.001	0.001	0.016	-0.008	0.000
$\text{Demand}^{6\times6}$	-0.004	-0.006	0.001	0.003	0.005	-0.001
$\text{Demand}^{\text{Idio}}$	0.000	0.000	-0.007	0.000	0.000	0.004

D.2 Results related to non-directional news indicators

In section 4.3, we construct three stock-level and three macro-level indicators of news arrival. This section provides further details about them.

Do news indicators explain return volatility? To gauge whether the news indicators are useful, we examine the relationship between them and weekly stock return volatility. We start with the stock-level news indicators.

To examine the IBES analyst update indicator, we sort the sample into five bins based on the indicator and then estimate regression:

$$|\text{Ret}_{i,t}| = \beta_1 + \sum_{\text{bin } b=2}^{5} \beta_b \cdot \mathbb{I}_{\text{IBES}_{i,t} \in \text{ bin } b} + \epsilon_{i,t} \tag{A10}$$

where $\text{Ret}_{i,t}$ is the weekly stock return over market and the independent variables are indicator variables for the five IBES indicator-sorted bins. We cluster standard errors by week and stock. The results are shown in the first column of Panel A in Table D.12. The baseline level of $|\text{Ret}_{i,t}|$ is 3.813% for the first bin and rises monotonically from the 3$^{\text{rd}}$ to the 5$^{\text{th}}$ bin. The weekly return magnitude is 1.778% higher in the top bin and the difference is statistically significant.

We conduct a similar exercise by sorting on Ravenpack and realized volatility (rvol) and find that they also explain return volatility, as reported in columns (2) and (3). The explanatory power of Ravenpack is similar to that of IBES and rvol is slightly lower. To gauge how related these stock-level indicators are, the first three columns of Panel B report the rank correlation between them. The IBES and Ravenpack measures have a sizeable correlation of 0.295, but the realized volatility measure is largely uncorrelated with the other two.

What about the macro-level news indicators? While the stock-level measures are all useful in explaining return volatility, the macro-news indicators are less successful. To gauge their usefulness, we first obtain weekly average values of stock-level $|\text{Ret}_{i,t}|$ into a single aggregate time series. We then estimate time-series regressions of it on bins sorted on the macro news indicators; the results are reported in columns (4) to (6) in Panel A of Table D.12. In column (4), we assign all weeks with one of the Savor-Wilson macroeconomic news releases into bin 2 and find that those weeks have 0.288% higher volatility. While the difference is statistically significant, the spread is not as large as the case with stock-level indicators. In column (5) which examines the ADS measure, we find that weeks in the top quintile have statistically significantly higher volatility than the baseline; the difference in the fourth bin is not statistically significant. In column (6), we do not find evidence that the Ravenpack macro news indicator explains return volatility. Overall, the macro news indicators are less successful in explaining return volatility. This might be related to the previous discovery that aggregate movements have a higher share of "discount rate news" than that at the stock-level (e.g. Campbell and Shiller, 1988; Vuolteenaho, 2002). Finally, columns (4) through (6) in Panel B find that the macro news indicators have weak correlations in the range of 0.022 to 0.083 amount themselves and weaker correlation with the stock-level news measures.

D.3 Sorting on contemporaneous realized volatility

In section 4.3, we use lagged realized volatility (rvol) to proxy for news arrival. This section explains why we do not use *contemporaneous* rvol: it is endogenous, and sorting on it leads to a mechanical bias in price multiplier estimation. This section explains this bias. We first provide an analytical framework and then present simulation and empirical results.

Mechanism for the bias. To begin with, contemporaneous rvol is correlated with the magnitude of realized weekly return ($|\text{Ret}_{i,t}|$), so sorting on contemporaneous rvol is similar to sorting on realized $|\text{Ret}_{i,t}|$. As an empirical verification, in Panel (a) of Figure D.9, we sort the sample into contemporaneous rvol deciles and plot the resulting average $|\text{Ret}_{i,t}|$. As expected, there is a monotonic relationship. From the bottom to top rvol decile, $|\text{Ret}_{i,t}|$ rises

Table D.12. News indicators and return volatility

Panel A presents regressions for examining the relationship between news indicators and the absolute value of weekly stock returns over market ($|\text{Ret}_{i,t}|$, in percent). In columns (1) through (3), we split the sample into quintiles using univariate sorts with the three stock-level news indicators, and then we regress $|\text{Ret}_{i,t}|$ on bin dummy variables. Standard errors are clustered by week and stock. In columns (4) through (6), we examine macro news indicators by regressing weekly average $|\text{Ret}_{i,t}|$ onto bins formed using univariate sorts with macro news indicators. We use "Ravenpack (S)" and "Ravenpack (M)" to denote the stock-level and macro-level Ravenpack news indicators, respectively. In column (4), the sample is split into two bins where bin = 2 indicates weeks with one of the three Savor and Wilson (2013) macro news releases. In columns (5) and (6), the sample is split into quintiles based on the absolute value of changes in the Aruoba-Diebold-Scotti (ADS) Business Conditions Index and Ravenpack macro news, respectively. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01. Panel B reports the rank correlation between the news indicators.

Panel A: relationship between news and return volatility

	Dependent variable: $\|\text{Ret}_{i,t}\|$, in percent						
	By stock-level news indicators				By macro news indicators		
Sorting variable	IBES	Ravenpack (S)	Realized vol		Savor-Wilson	ADS	Ravenpack (M)
	(1)	(2)	(3)		(4)	(5)	(6)
Intercept	3.813***	3.816***	3.705***	Intercept	3.667***	3.736***	3.792***
	(0.059)	(0.062)	(0.053)		(0.063)	(0.095)	(0.095)
Bin=2	−0.005	0.006	0.221***	Bin=2	0.288***	−0.060	0.089
	(0.009)	(0.012)	(0.027)		(0.085)	(0.134)	(0.135)
Bin=3	0.110***	0.135***	0.460***	Bin=3		0.048	0.049
	(0.014)	(0.018)	(0.044)			(0.134)	(0.135)
Bin=4	0.313***	0.309***	0.767***	Bin=4		0.179	0.001
	(0.027)	(0.023)	(0.070)			(0.134)	(0.135)
Bin=5	1.778***	1.732***	1.285***	Bin=5		0.272**	0.019
	(0.049)	(0.043)	(0.142)			(0.134)	(0.135)
Obs	2,270,134	2,270,134	2,270,134	Obs	1,148	1,148	1,148
Adj R^2	0.020	0.018	0.009	Adj R^2	0.009	0.004	-0.003

Panel B: rank correlation between news indicators

	IBES	Ravenpack (S)	Realized vol		Savor-Wilson	ADS	Ravenpack (M)
	(1)	(2)	(3)		(4)	(5)	(6)
IBES	1	0.295	0.044	IBES	0.028	0.006	0.024
Ravenpack (S)	0.295	1	0.022	Ravenpack (S)	0.022	0.007	0.027
Realized vol	0.044	0.022	1	Realized vol	0.028	0.068	-0.020
					Savor-Wilson	ADS	Ravenpack (M)
				Savor-Wilson	1	0.049	0.083
				ADS	0.049	1	0.022
				Ravenpack (M)	0.083	0.022	1

from 3.01% to 6.33%.

Sorting on the magnitude of realized return leads to a bias in multiplier estimation.

Consider the following simple data-generating process (DGP):

$$\text{Ret}_{i,t} = M \cdot \text{Demand}_{i,t} + \epsilon_{i,t} \tag{A11}$$

where the true price multiplier M is a constant. The two shocks, $\text{Demand}_i \sim N(0, \sigma^2_{\text{Demand}})$ and $\epsilon_i \sim N(0, \sigma^2_\epsilon)$, are independent from each other.

What does sorting on realized $|\text{Ret}_{i,t}|$ achieve? Well, $|\text{Ret}_{i,t}|$ is high (low) when the *realizations* of $\text{Demand}_{i,t}$ and $\epsilon_{i,t}$ are positively (negatively) correlated. Therefore, if we restrict attention to periods with high (low) realized $|\text{Ret}_{i,t}|$, the regression estimates of M will be biased up (down). While our simple DGP only incorporates stock-level variation, similar concerns exist for studying price multipliers at more aggregate levels when sorting on realized return magnitudes.

Simulation. We now verify this bias by simulating from equation (A11) using parameters that are in the ballpark of actual data. We pick $M = 1.732$ based on Table 3, and we choose $\sigma_{\text{Demand}} = 0.64\%$ and $\sigma_\epsilon = 5.8\%$ so the volatility of $\text{Demand}_{i,t}$ and $\text{Ret}_{i,t}$ will match the real data. We simulate 10,000,000 observations.

We sort the simulated sample into 50 bins based on $|\text{Ret}_{i,t}|$. In Panel (b) of Figure D.9, we plot the correlation between $\text{Demand}_{i,t}$ and $\epsilon_{i,t}$ against the average $|\text{Ret}_{i,t}|$ by bin. As anticipated, even though $\text{Demand}_{i,t}$ and $\epsilon_{i,t}$ are independent in the population, subsamples with higher $|\text{Ret}_{i,t}|$ have higher in-sample correlations between $\text{Demand}_{i,t}$ and $\epsilon_{i,t}$.

We then proceed to estimate price multipliers by regressing $\text{Ret}_{i,t}$ on $\text{Demand}_{i,t}$ in each bin. In Panel (c), we plot the estimated multipliers against $|\text{Ret}_{i,t}|$ over a range similar to that of the real data. As predicted, even though the true price multiplier is homogeneous by construction, subsamples with higher $|\text{Ret}_{i,t}|$ exhibit spurious heterogeneity. To get a sense of the range of spurious multiplier variation, we use blue dashed lines to mark the range of $|\text{Ret}_{i,t}|$ spanned by the ten rvol deciles in actual data (Panel (a)). When considering the variation between the blue lines, we find that the estimated multipliers vary from approximately

0.4 to 2.

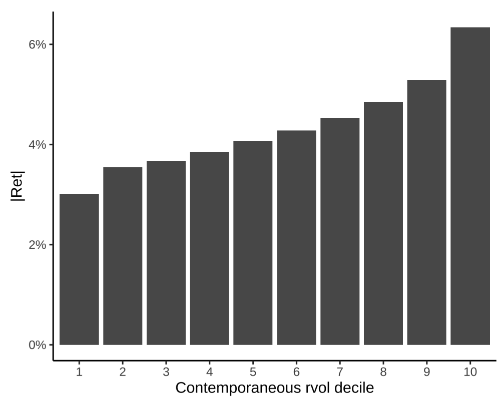
Detecting this bias in real data. We now show evidence that such bias also shows up in the real data. In Panel (d), we sort our weekly data into contemporaneous rvol deciles and then estimate price multipliers for all demand components for each decile. As predicted, for all demand components, multipliers vary monotonically with rvol. In fact, the estimated stock-level multiplier varies monotonically from 0.8 to 2 which is similar to that in the simulated results. This suggests that sorting on contemporaneous rvol leads to a similar degree of mechanical multiplier variation that is similar to that in simulated data.

To avoid this mechanical bias, in section 4.3, we sort the sample using one-week-lagged rvol instead of contemporaneous rvol in the paper. Panel (e) confirms that lagged rvol is useful for forecasting contemporaneous rvol. We sort stocks cross-sectionally using lagged rvol into deciles and plot average current rvol against lagged rvol. While the results do not lie on the 45% degree line, lagged rvol does help predict current rvol. When we sort on lagged rvol, we no longer find clear variation in price multipliers across subsamples, a finding that is plotted in Panel (f) of Figure D.9.

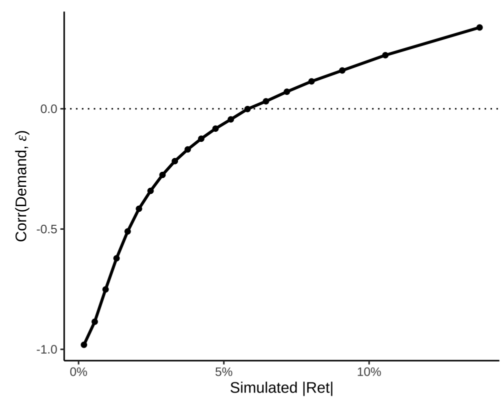
Figure D.9. Estimating price multipliers on rvol-based subsamples

In Panel (a), we sort our weekly sample using contemporaneous realized volatility (rvol) into deciles and plot the average realized absolute value of weekly returns over market (|Ret|). We then simulate demand and return data under the null hypothesis of a homogeneous price multiplier according to equation (A11) and sort the simulated data into bins using realized |Ret|. Panel (b) plots the correlation between demand and ϵ for each bin in the simulated data. Panel (c) plots the price multiplier estimates for each bin in the simulated data; the dashed blue lines mark the range of |Ret| spanned by the rvol deciles in Panel (a). We then return to the real data, sort it into deciles by contemporaneous (lagged) rvol, and plot price multiplier estimates by decile in Panel (d) (and Panel (f)). The error bars represent 95% confidence intervals. In Panel (e), we sort the sample into deciles by lagged rvol and plot the average contemporaneous rvol against lagged rvol by decile.

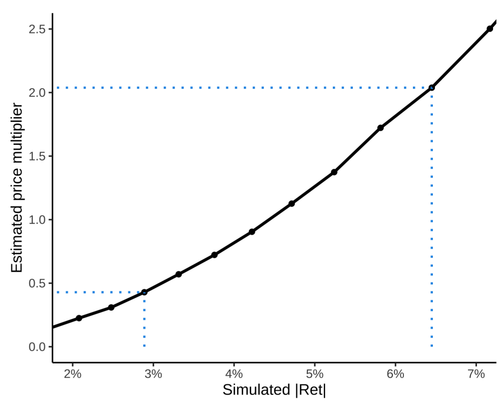
(a) |Ret| by contemporaneous rvol



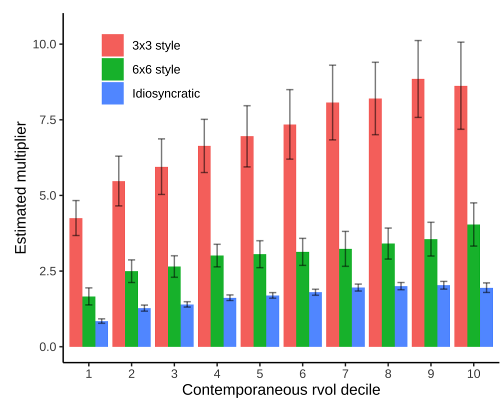
(b) Simulated: Cor(Demand, ϵ) by |Ret|



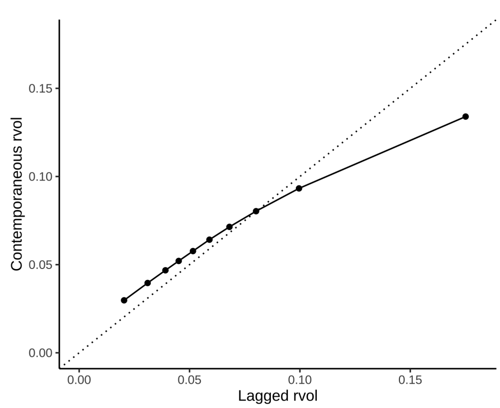
(c) Simulated: multiplier by |Ret|



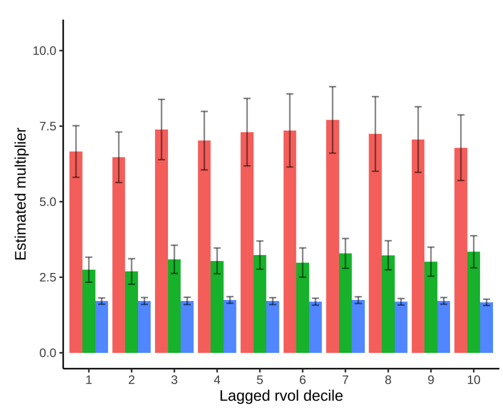
(d) Multipliers by contemporaneous rvol



(e) Lagged and contemporaneous rvol



(f) Multipliers by lagged rvol



E Theoretical mechanisms

This section explores existing models of price impact discussed in section 5 in more detail. Appendix E.1 discusses the demand-system approach following Koijen and Yogo (2019). Appendix E.2 solves and calibrates the risk-based model in section 5.2.

E.1 Koijen and Yogo (2019) demand systems

In equilibrium models, price multipliers are reciprocals of (aggregate) investor demand elasticities. Thus, to generate the findings in our paper, investors need to exhibit lower demand elasticities at more aggregate levels. However, models following Koijen and Yogo (2019) (henceforth KY) do not have this feature.

The model. We introduce a simplified version of the KY model for the convenience of readers. Consider a one-period model with $n \in \{0, 1, ..., N\}$ assets, normalized to one share outstanding for each. Assets n =1 to N are stocks and $n = 0$ is an "outside asset" which represents investment opportunities outside of the stock market. Let $p(n)$ denote the log price of stock n. Because each stock has one share outstanding, $p(n)$ is also the log market capitalization. To discuss demand elasticities at stock versus style levels, assume that stocks $\{1, 2, ..., N\}$ are partitioned into disjoint style portfolios $s \in \{1, 2, ..., S\}$ with $S < N$.

Each investor $i \in \{0, 1, ..., I\}$ has wealth A_i which is exogenously given in the baseline model. Each investor i chooses a portfolio from her exogenously determined investment universe $\mathcal{I}_i \subset \{0, 1, ..., N\}$ to which the outside asset $n = 0$ is always included. Investor i's portfolio weights are given by:

$$w_i(n) = \frac{\delta_i(n)}{\sum_{m \in \mathcal{I}_i} \delta_i(m)} \qquad \forall n \in \mathcal{I}_i \tag{A12}$$
$$\text{where } \delta_i(n) = e^{\beta_i p(n) + v_i(n)}$$

where β_i is an investor-specific constant ($\beta_{i,0}$ in equation (10)) that governs her stock-level demand elasticity $(1-\beta_i)$. Assume that $\beta_i \leq 1$ so demand curves are weakly downward sloping. $v_i(n)$ denotes additional drivers of stock n demand that is driven by preferences for non-price characteristics and latent demand. We lump all those terms into $v_i(n)$ as they are unrelated to how demand responds to prices.

The aggregate demand for stock n in log number of shares is given by:

$$q(n) = \log\left(\sum_{i:n\in\mathcal{I}_i} A_i \cdot w_i(n)\right) - p(n) \tag{A13}$$

E.1.1 Baseline results with homogeneous investors

We start with the simpler case with a single representative investor. We omit investor index i and simplify (A13) to

$$q(n) = \log A + \log w(n) - p(n). \tag{A14}$$

To study demand elasticities at the stock and style levels, we consider a small shock dp to the log prices of either a single stock n or all stocks in a style s. Differentiating (A14) gives a decomposition of demand elasticities:

$$-\frac{dq(n)}{dp} = 1 - \underbrace{\frac{d\log A}{dp}}_{\text{wealth effect}} - \underbrace{\frac{d\log w(n)}{dp}}_{\text{substitution effect}}, \tag{A15}$$

and we now discuss the impact of these two terms in turn.

1. **Substitution effect.** Based on (A12), we have:

$$
\begin{aligned}
-\frac{d\log w(n)}{dp} &= -\frac{d}{dp}\log\left(\frac{\delta(n)}{\sum_m \delta(m)}\right) \\
&= -\frac{\sum_m \delta(m)}{\delta(n)} \cdot \frac{(\sum_m \delta(m)) \cdot [d\delta(n)/dp] - \delta(n) \cdot [d(\sum_m \delta(m))/dp]}{[\sum_m \delta(m)]^2} \\
&= \begin{cases} -(1-w(n)) \cdot \beta & \text{for stock-level shock to } n \\ -(1-w(s)) \cdot \beta & \text{for style-level shock to } s \end{cases}
\end{aligned}
$$

where $w(s) = \sum_{m \in s} w(m)$ is the portfolio weight in style s. On average, $w(s) > w(n)$, so this channel generates *higher* style-level demand elasticity, which is opposite to our empirical findings. The key mechanism for the difference is that, when the price of an entire style portfolio changes, it creates a larger effect on the denominator in equation (A12), a point that is also noted in Van der Beck (2022).

2. **Wealth effect.**

   - The standard KY model assumes exogenous wealth so this channel is mute.
   - If one endogenizes the effect of price shocks on wealth (e.g. Darmouni, Siani, and Xiao, 2022), then this goes in the opposite direction to the substitution effect, because $w(s)$ is typically larger than $w(n)$. Specifically, we have

$$
-\frac{d\log A}{dp} = \begin{cases} -w(n) & \text{if stock-level shock to } n \\ -w(s) & \text{if style-level shock to } s \end{cases}
$$

   which should be intuitive as style-level shocks have a larger impact on wealth.

We now wrap up the comparison between stock- and style-level demand elasticities. If wealth is exogenous as in the original KY model, then we get higher style-level demand elasticities which runs counter to the empirical results:

$$\begin{aligned}\text{DE}_{\text{stock}} &= 1 - \left(\frac{d\log w(n)}{dp}\right)_{\text{stock level}} - \left(\frac{d\log(A)}{dp}\right)_{\text{stock level}} \\ &= 1-(1-w(n))\cdot\beta. \quad \text{Similarly,} \\ \text{DE}_{\text{style}} &= 1-(1-w(s))\cdot\beta > \text{DE}_{\text{stock}}\end{aligned}$$

If wealth is endogenous, then we have the desired direction with lower style-level demand elasticity:

$$\begin{aligned}\text{DE}_{\text{stock}} &= 1-(1-w(n))\cdot\beta - w(n) \\ &= (1-w(n))\cdot(1-\beta). \quad \text{Similarly,} \\ \text{DE}_{\text{style}} &= (1-w(s))\cdot(1-\beta) < \text{DE}_{\text{stock}} \text{ as long as } w(s) > w(n)\end{aligned}$$

In this version with endogenous wealth, the model predicts the following ratio between the two demand elasticities:

$$\text{DE}_{\text{style}} = \frac{1-w(s)}{1-w(n)} \times \text{DE}_{\text{stock}} \approx (1-w(s))\cdot\text{DE}_{\text{stock}} \tag{A16}$$

This is in the right direction, but this mechanism cannot generate quantitatively large differences. Consider the definition of a style as one of the 3×3 size-BM portfolios as in our paper. The average $w(s)$ for a style is thus $1/9$, so this predicts that the style-level price multiplier is $1/(1-1/9) = 1.125$ times the stock-level one — only mildly higher. In contrast, according to estimates in Table 3, the ratio is $6.976/1.723 \approx 4$ in the data.

As we will explain in Appendix E.1.2, to match data using this channel, we need investor portfolios to be extremely segmented at the style-level so that $w_i(s) \approx 3/4$. That is, most investors need to restrict their portfolio to a single style and not invest in other styles.

E.1.2 Results with investor heterogeneity

In practice, investors hold heterogeneous portfolios, and a major goal of Koijen and Yogo (2019) is to take into account the rich heterogeneity across 13F institutions. In principle, if investor portfolios are highly segmented across styles, this can generate larger style-level multipliers. We explain the logic in this section. However, when calibrated to actual institutional holdings data, the effect is still quantitatively small.

How does portfolio segmentation impact price multipliers? For intuition, consider the extreme case where the portfolio of every investor is exogenously constrained to a single style. Then, style-level demand elasticity would be zero and the associated price multiplier would rise to infinity.[9] This is extreme, but it shows that more style-level specialization can lead to lower aggregate style-level demand elasticities.

We now derive the relationship between the *degree* of portfolio segmentation and price multipliers to assess the *quantitative* impact of this channel. Appendix E.1.3 provides detailed derivations and we summarize the findings here. Rewriting equation (A16) in the case with heterogeneous agents says that for each investor i, her demand elasticity for style s equals her stock-level demand elasticity multiplied by an attenuation factor:

$$\mathrm{DE}_{i,\text{style}}(s) \approx \underbrace{(\mathbf{1} - \mathbf{w_i}(\mathbf{s}))}_{\text{attenuation factor}} \cdot \mathrm{DE}_{i,\text{stock}}. \tag{A17}$$

where the attenuation factor $(1 - w_i(s))$ is small if $w_i(s)$ — her portfolio weight in style s — is large. In other words, the more concentrated the investor's portfolio in style s, the lower her style-level demand elasticity relative to the stock-level one. This nests the earlier discussion as a special case: if the investor only holds stocks in style s, then $w_i(s) = 0$, and her style-level demand elasticity is zero.

[9] The mechanism is simple. If only the price of a single stock n in style s declined, then investors in style s can increase their demand in n and reduce their demand in other stocks in s which limits the price impact on stock n. However, if the price of *all* stocks in style s declined equally, then investors in style s have no reason to adjust their portfolios, and investors in other styles will also not adjust because they are constrained from investing in style s. As a consequence, the aggregate style-level demand elasticity is zero.

It turns out that a relationship similar to (A17) holds when considering *aggregate* demand elasticities, except we need to take a weighted average over the dollar value of investor holdings:

$$\mathrm{DE}^{\mathrm{agg}}_{\mathrm{stock}} \approx \frac{\sum_i A_i \cdot w_i(s) \cdot \mathrm{DE}_{i,\mathrm{stock}}}{\sum_i A_i \cdot w_i(s)} \tag{A18}$$

$$\mathrm{DE}^{\mathrm{agg}}_{\mathrm{style}}(s) \approx \frac{\sum_i A_i \cdot w_i(s) \cdot (\mathbf{1} - \mathbf{w_i}(\mathbf{s})) \cdot \mathrm{DE}_{i,\mathrm{stock}}}{\sum_i A_i \cdot w_i(s)}. \tag{A19}$$

where $A_i \cdot w_i(s)$ is the dollar value of style s held by investor i.

Calibration shows that the effect is small. In practice, how concentrated are investor holdings at the style level? To get a sense, we follow Koijen and Yogo (2019) to examine 13F institutional holdings in the U.S. stock market. We use Q4 of 2017 which is the last period in their sample. When restricting attention to the sample of stocks considered in our paper, around 70% of market value is held by 13F institutions. We follow KY to construct a "household" sector that holds the remaining 30% of stocks.

We study investor portfolio concentration in 3×3 size-BM sorted style portfolios and report the results in Table E.13. The first nine rows report results separately for each style and the last row reports the average across styles. To establish a benchmark, we first consider the case with a representative investor. Column (1) reports the style market weights $w(s)$ which are the portfolio weights for the representative investor, and column (2) reports $1/(1 - w(s))$ — the implied ratio of style- and stock-level multipliers. The average implied ratio is 1.182 across styles, only slightly larger than 1, so it is too small to explain the empirical findings of a ratio around 4.

Column (3) reports the holdings-weighted average ratio of $1/(1 - w(s))$ across all 13F institutions and the household sector. Concretely, for each style s and each investor i that holds some stocks in the style, we compute $1/(1 - w_i(s))$ and winsorize it at 0.01% and 99.99% levels to avoid extreme values. We then compute the holdings value-weighted average

Table E.13. Model-implied ratio between style- and stock-level multipliers

This table reports the implied ratio between style- and stock-level price multipliers in KY-style models. The results are calibrated using data as of the last quarter of 2017. Each of the first nine rows represents one of the 3×3 size/BM-sorted portfolios and the last row reports the average across styles. We first consider the case with a representative investor who holds the whole market. Column (1) reports the market weight of each style portfolio which is also the portfolio weight of the representative investor; column (2) reports $1/(1-w(s))$ which is the implied ratio between the style- and stock-level price multipliers. We then consider the case with heterogeneous investors and report the holdings-weighted average of $1/(1-w(s))$ calibrated using 13F institution holdings in the next two columns; columns (3) and (4) report the cases with or without the residual household (HH) sector, respectively. We then estimate the holdings-weighted average of $1/(1-w(s))$ using mutual funds and report the results in columns (5) through (7). Column (5) uses all mutual funds, and the results when restricting to mutual funds with turnover $\geq 100\%$ or $\geq 200\%$ are reported in columns (6) and (7).

Style s		$w(s)$	$1/(1-w(s))$					
Size	BM	Market	Market	13F institutions		Mutual funds, with turnover		
				With HH	No HH	$\geq 0\%$	$\geq 100\%$	$\geq 200\%$
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Small	Growth	0.013	1.013	1.014	1.014	1.033	1.051	1.087
Small	Blend	0.013	1.014	1.014	1.016	1.041	1.055	1.072
Small	Value	0.011	1.011	1.011	1.012	1.030	1.038	1.039
Mid	Growth	0.043	1.045	1.050	1.058	1.092	1.136	1.131
Mid	Blend	0.032	1.033	1.034	1.039	1.054	1.075	1.075
Mid	Value	0.024	1.024	1.025	1.028	1.040	1.055	1.054
Large	Growth	0.518	2.075	2.230	2.288	2.773	2.850	2.619
Large	Blend	0.213	1.271	1.272	1.256	1.272	1.278	1.305
Large	Value	0.133	1.154	1.162	1.164	1.190	1.179	1.185
Average		0.111	1.182	1.201	1.208	1.281	1.302	1.285

of $1/(1-w_i(s))$ across investors and report the result in column (3). As reported in the last row, the average implied ratio is 1.201 which is only marginally higher than the case with a representative investor (column (2)). In column (4), we report the results without considering the household sector and find similar conclusions.

One may be concerned that 13F holdings may be too coarse as it is reported at the management company level. For instance, suppose a management company operates nine separate funds, each of which invests exclusively in one of the 3×3 styles. If the holdings are reported at the fund level, we would correctly conclude that all funds are fully style-concentrated, but when reported at the management company level, we would wrongly conclude that the holdings are well diversified across styles.

To gauge the importance of this concern, we also redo the exercise using Thomson Reuters

S12 data which reports holdings by mutual funds. The results are reported in column (5). While this does increase the ratio slightly, the average ratio is still only at 1.281 which is much lower than that needed to explain our findings. One may be further concerned that our exercise involves too many passive funds which tend to hold diversified index portfolios, while the active funds — who matter more for liquidity provision — may be more style-concentrated. To evaluate this concern, we further restrict attention to mutual funds with annual turnover above 100% or 200% and report the results in columns (6) and (7). This does not lead to a significant change in results, indicating that active mutual funds also tend to be diversified across styles. Overall, we conclude that when considering realistic data on investor holdings, KY-style models with heterogeneous investors cannot generate a large difference between style- and stock-level price multipliers.

E.1.3 Model with investor heterogeneity: derivation details

Let $Q_i(n)$ denote the number of stock n shares demanded by investor i. The aggregate stock-level demand elasticity is simply the holdings-weighted average of investor-level stock demand elasticities:

$$\begin{aligned}
\mathrm{DE}^{\mathrm{agg}}_{\mathrm{stock}}(n) &= -\frac{d\log(\sum_i Q_i(n))}{dp} \\
&= \frac{1}{\sum_i Q_i(n)}\left(\sum_{i:n\in\mathcal{I}_i} -\frac{dQ_i(n)}{dp}\right) \\
&\overset{\mathrm{DE}_{i,\mathrm{stock}}=-\frac{dq_i(n)}{dp}}{=} \frac{\sum_{i:n\in\mathcal{I}_i} Q_i(n)\cdot \mathrm{DE}_{i,\mathrm{stock}}}{\sum_i Q_i(n)} \\
&= \frac{\sum_{i:n\in\mathcal{I}_i} A_i\cdot w_i(n)\cdot \mathrm{DE}_{i,\mathrm{stock}}}{\sum_{i:n\in\mathcal{I}_i} A_i\cdot w_i(n)}. \qquad \text{(A20)}
\end{aligned}$$

Following the same derivation steps, for each style s, we have:

$$\begin{aligned}\text{DE}^{\text{agg}}_{\text{style}}(s) &= \frac{\sum_{i:s\in\mathcal{I}_i} A_i \cdot w_i(s)\cdot \text{DE}_{i,\text{style}}(s)}{\sum_{i:s\in\mathcal{I}_i} A_i \cdot w_i(s)} \\ &\overset{\text{(A16)}}{\approx} \frac{\sum_{i:s\in\mathcal{I}_i} A_i \cdot w_i(s)\cdot (\mathbf{1}-\mathbf{w_i(s)})\cdot \text{DE}_{i,\text{stock}}}{\sum_{i:s\in\mathcal{I}_i} A_i \cdot w_i(s)}. \end{aligned} \tag{A21}$$

which ends the derivation for equation (A19). Finally, we computed the value-weighted average $\text{DE}^{\text{agg}}_{\text{stock}}(n)$ for all stocks in the same style:

$$\begin{aligned}\text{DE}^{\text{agg}}_{\text{stock}} &\equiv \frac{\sum_{n\in s} P(n)\cdot \text{DE}^{\text{agg}}_{\text{stock}}(n)}{\sum_{n\in s} P(n)} \\ &\overset{\text{(A20)}}{=} \frac{\sum_{n\in s}\sum_{i:n\in\mathcal{I}_i} A_i \cdot w_i(n)\cdot \text{DE}_{i,\text{stock}}}{\sum_{n\in s}\sum_{i:n\in\mathcal{I}_i} A_i w_i(n)} \\ &= \frac{\sum_{i:s\in\mathcal{I}_i} A_i \cdot w_i(s)\cdot \text{DE}_{i,\text{stock}}}{\sum_{i:s\in\mathcal{I}_i} A_i \cdot w_i(s)} \end{aligned} \tag{A22}$$

which ends the derivation for equation (A18). Finally, note that equation (A22) is similar to (A21) except the $(1 - w_i(s))$ attenuation factor. Thus, just like in the representative agent case, the ratio between style- and stock-level demand elasticities in a heterogeneous KY model is also given by a factor of $1-w(s)$. The difference is, that instead of the factor for the representative investor, it is now the value-weighted average factor across all investors.

E.2 Solving and calibrating the model in section 5.2

Solving for equilibrium. The expected returns and covariances are given by:

$$E(R_i) = a_i - P_i \qquad \forall i = 1, ..., N$$

$$Cov(R_i, R_j) = \begin{cases} \sigma^2_{\text{market}} + \sigma^2_{\text{style}} + \sigma^2_{\text{idio}} & \text{if } i = j \\ \sigma^2_{\text{market}} + \sigma^2_{\text{style}} & \text{if } i \neq j \text{ but } s(i) = s(j) \\ \sigma^2_{\text{market}} & \text{if } s(i) \neq s(j) \end{cases}$$

To get (17), note that CARA-normal optimization is equivalent to mean-variance optimization. Let Σ denote the return covariance matrix and let x denote the vector $(x_1, ..., x_N)'$ for any variable $\{x_i\}_i$. Then, the aggregate investor demand for stock shares is given by:

$$D = \frac{1}{\gamma}\Sigma \cdot E(R) = \frac{1}{\gamma}\Sigma \cdot (a - P) \tag{A23}$$

We then solve for prices by equaling supply and demand:

$$\begin{aligned} Z = D &= \frac{1}{\gamma}\Sigma \cdot (a - P) \\ \Rightarrow P &= a - \gamma \Sigma Z \end{aligned} \tag{A24}$$

For each stock i, (A24) is:

$$\begin{aligned} P_i &= a_i - \gamma \sum_{j=1}^{N} Cov(R_i, R_j) \cdot Z_j \\ &= a_i - \gamma \left[\sigma^2_{\text{market}} \cdot \sum_{j=1}^{N} Z_j + \sigma^2_{\text{style}} \cdot \sum_{j \in s(i)} Z_j + \sigma_{\text{idio}} \cdot Z_i \right] \\ &= a_i - \gamma \left[\underbrace{\left(N \cdot \sigma^2_{\text{market}}\right) \cdot Z^{\text{market}}}_{\text{market-level risk}} + \underbrace{\left(N_s \cdot \sigma^2_{\text{style}}\right) \cdot Z^{\text{style}}_{s(i)}}_{\text{style-level risk}} + \underbrace{\sigma^2_{\text{idio}} \cdot Z_i}_{\text{idiosyncratic risk}} \right] \end{aligned} \tag{A25}$$

where we defined $Z^{\text{market}} = \frac{1}{N}\sum_{j=1}^{N} Z_j$ and $Z^{\text{style}}_s = \frac{1}{N_s}\sum_{j \in s} Z_j$. Rearranging (A25) gives (17) in the paper.

Price multipliers. We now compare style- and stock-level price multipliers. Equation (A25) implies that, for any stock i, the partial derivative of price to supply shocks are:

$$-\frac{\partial P_i}{\partial Z^{\text{style}}_s} = N_s \cdot \sigma^2_{\text{style}} \text{ , and} \tag{A26}$$

$$-\frac{\partial P_i}{\partial Z_i} = \sigma^2_{\text{idio}}. \tag{A27}$$

Then, the style-level price multiplier is given by:

$$
\begin{aligned}
M^{\text{style}} &\equiv -\frac{1}{N_s}\sum_{i\in s}\frac{\partial P_i}{\partial Z_s^{\text{style}}}\cdot\frac{Z_s^{\text{style}}}{P_i} \\
&\overset{\text{plug in (A26)}}{=} \sigma_{\text{style}}^2\cdot\sum_{i\in s}\frac{Z_s^{\text{style}}}{P_i} \\
&= N_s\sigma_{\text{style}}^2\cdot\frac{Z_s^{\text{style}}}{P_s^{\text{style}}}
\end{aligned}
\tag{A28}
$$

where $P_s^{\text{style}} \equiv \left(\sum_{i\in s} 1/P_i\right)^{-1}$. The average stock-level multiplier in style s is given by:

$$
\begin{aligned}
M^{\text{idio}} &= -\frac{1}{N_s}\sum_{i\in s}\frac{\partial P_i}{\partial Z_i}\cdot\frac{Z_i}{P_i} \\
&\overset{\text{plug in (A27)}}{=} \frac{1}{N_s}\sum_{i\in s}\sigma_{\text{idio}}^2\cdot\frac{Z_i}{P_i} \\
&= \sigma_{\text{idio}}^2\cdot\frac{Z_s^{\text{style}}}{P_s^{\text{style}}}
\end{aligned}
\tag{A29}
$$

Dividing equation (A28) by (A29) gives the scaling relationship

$$
\frac{M^{\text{style}}}{M^{\text{idio}}} = \frac{N\sigma_{\text{style}}^2}{\sigma_{\text{idio}}^2}
\tag{A30}
$$

which is equation (18) in the paper.

Calibration. We now calibrate this model to data moments. We set a_i such that P_i is normalized to 1 for all stocks i so that the variance of payoffs is the same as the variance of returns. We take the "styles" in this model to be the 3×3 portfolios in our data. Because there are $S = 9$ styles and an average of $N = 2,507$ stocks, each style has an average of $N_s = N/S \approx 279$ stocks.

We adjust parameters to match return volatilities. Note that in this model, stock, style, and market gross returns are given by (recall that prices are all normalized to 1):

$$
\begin{aligned}
R_i &= F_m + F_{s(i)} + \epsilon_i && \forall i = 1, ..., N \\
R_s^{\text{style}} &= \frac{1}{N_s} \sum_{i \in s} R_i = F_m + F_s + \frac{1}{N_s} \sum_{i \in s} \epsilon_i && \forall s = 1, ..., S \\
R^{\text{market}} &= \frac{1}{N} \sum_{i=1}^{N} R_i = F_m + \frac{1}{S} \sum_{s=1}^{S} F_s + \frac{1}{N} \sum_{i=1}^{N} \epsilon_i
\end{aligned}
$$

Table 1 contains return volatilities of market-hedged style returns and style-hedged stock returns. for the former, its variance in the model is:

$$
\begin{aligned}
Var(R_s^{\text{style}} - R^{\text{market}}) &= Var\left(\frac{S-1}{S} F_s - \frac{1}{S} \sum_{s' \neq s} F_{s'} + \frac{S-1}{N} \sum_{i \in s} \epsilon_i - \frac{1}{N} \sum_{j \neq s} \epsilon_j \right) \\
&= \left[\frac{(S-1)^2}{S^2} + \frac{S-1}{S^2} \right] \sigma_{\text{style}}^2 + \left[\frac{(S-1)^2}{NS} + \frac{(S-1)}{NS} \right] \sigma_{\text{idio}}^2 \\
&= \frac{S-1}{S} \sigma_{\text{style}}^2 + \frac{S-1}{N} \cdot \sigma_{\text{idio}}^2
\end{aligned} \tag{A31}
$$

Similarly, for the style-hedge stock return, we have:

$$
\begin{aligned}
Var(R_i - R_{s(i)}^{\text{style}}) &= Var\left(\epsilon_i - \frac{1}{N_s} \sum_{j \in s(i)} \epsilon_j \right) \\
&= Var\left(\frac{N_s - 1}{N_s} \epsilon_i - \frac{1}{N_s} \sum_{j \in \{s(i)\ \ i\}} \epsilon_j \right) \\
&= \left(\frac{(N_s - 1)^2}{N_s^2} + \frac{N_s - 1}{N_s^2} \right) \sigma_{\text{idio}}^2 \\
&= \left(1 - \frac{1}{N_s} \right) \sigma_{\text{idio}}^2
\end{aligned} \tag{A32}
$$

We choose σ_{style} and σ_{idio} so that these return variances match data. According to Table

1, the annualized variance of these return components are:

$$Var(R_s^{\text{style}} - R^{\text{market}}) = 52 \times (1.50\%)^2 \approx 0.0117$$
$$Var(R_i - R_{s(i)}^{\text{style}}) = 52 \times (5.92\%)^2 \approx 0.1822$$

Using (A31) and (A32) to match the variances above gives $\sigma_{\text{style}} \approx 11.2\%$ and $\sigma_{\text{idio}} \approx 42.8\%$.

Plugging these into (A30) gives:

$$\frac{M_{\text{style}}}{M_{\text{stock}}} = N_s \cdot \frac{\sigma_{\text{style}}^2}{\sigma_{\text{idio}}^2} \overset{\text{plug in calibration}}{\approx} 279 \times \frac{11.2\%^2}{42.8\%^2} \approx 19$$

F Dynamics of demand and price impact

In the main paper, we use contemporaneous regressions to estimate price effects. This section studies the dynamics of demand and price impact. Appendix F.1 shows that a small fraction of demand can be predicted and that price impacts almost entirely arise from unexpected demand variation. Appendix F.2 reconciles the price multiplier estimates by observational frequencies.

F.1 Expected versus surprise demand

Models such as Duffie (2010) and Gabaix and Koijen (2022) predict that price impact parameters of surprise demand components should be much larger than that of expected components. When estimating price multipliers in the main paper at weekly or slower frequencies, we do not differentiate between expected versus surprise demand variation. This is because most demand variation is unexpected at these frequencies. However, when separately examining expected and surprise demand, we do verify that the two command rather different price multipliers, a point that we demonstrate in this section.

Decomposing expected and surprise demand. For each demand component lev $\in \{3 \times 3 \text{ style}, 6 \times 6 \text{ style}, \text{Idiosyncratic}\}$, we estimate panel regressions with H lags of demand and returns:

$$\text{Demand}_{i,t}^{\text{lev}} = \sum_{h=1}^{H} b_h \cdot \text{Demand}_{i,t-h}^{\text{lev}} + \sum_{h=1}^{H} c_h \cdot \text{Ret}_{i,t-h}^{\text{lev}} + \sum_{t=1}^{T} \eta_t^{\text{lev}} + \epsilon_{i,t}^{\text{lev}} \tag{A33}$$

where $\{\eta_t^{\text{lev}}\}_{t=1}^{T}$ are time fixed effects. To obtain out-of-sample estimates, for each year $y = 2000, ..., 2022$, we use data from the year 1993 to $y - 1$ to estimate the relationship. The first quarter of the sample (1993 to 1999) is used as the "burn-in" period.

How many lags H should we include? Panel (a) of Figure F.10 plots the R^2 of demand prediction as we vary H. For the 3×3 and $6\times$ levels, maximum R^2 is achieved by $H = 8$ lags. While explanatory power does increase further with more lags for the idiosyncratic component, the marginal benefit of including more lags also tapers off. Therefore, for simplicity, we choose $H = 8$ lags for demand prediction at all levels. Based on this demand prediction model, the average R^2 for 3×3, 6×6, and idiosyncratic levels are 6.84%, 4.77%, and 9.85%, respectively, indicating that most demand variations are unpredictable. Panels (b) to (d) show that our demand predictions are well calibrated: realized average demand is aligned with model predictions.

Price multipliers of expected and surprise demand. We then use the predicted demand variation to decompose demand at each level into predicted and surprise components,

$$\text{Demand}_{i,t}^{\text{lev}} = \text{Demand}_{i,t}^{\text{expected, lev}} + \text{Demand}_{i,t}^{\text{surprise, lev}},$$

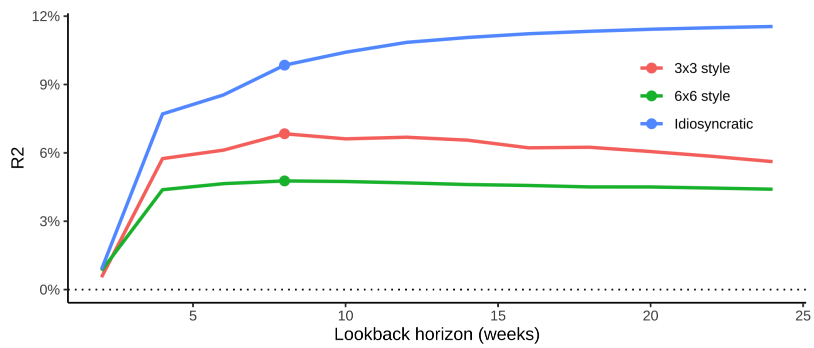
and estimate their price multipliers separately. The results are reported in Table F.14.

As shown in column (1) which is estimated using the full sample, the expected demand components have almost no price impact. The price multipliers of 3×3 and 6×6 levels are statistically indistinguishable from zero for expected demand. The multiplier associated

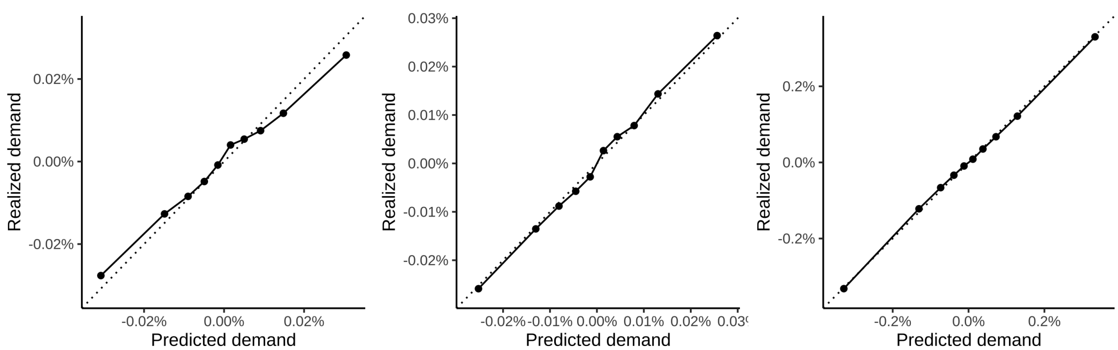
Figure F.10. Predicting demand

For each level in $\{3 \times 3, 6 \times 6, \text{idiosyncratic}\}$, we use H lags of weekly demand and return to predict demand out-of-sample using expanding windows for relationship estimation. Panel (a) plots the out-of-sample R^2 when varying H; the dots mark the specification using $H = 8$ lags. Panels (b) through (d) examine whether the demand prediction using $H = 8$ weeks is well calibrated. Specifically, for demand at each level, we sort by the predicted demand into ten deciles and plot the average realized demand against the predicted demand. The dashes lines are 45% degree lines.

(a) Out-of-sample R^2 when varying the number of lags



(b) Calibration: 3×3 level (c) Calibration: 6×6 level (d) Calibration: Idiosyncratic



with the expected idiosyncratic demand is statistically significant at the 5% level, but its magnitude is an order of magnitude smaller than that of the surprise component. Column (2) verifies that the finding is not sensitive to omitting the surprise components in the regressions. It is also worth noting that the within-period R^2 in column (2) drops to almost zero, indicating that almost all the demand-driven return variation arises from unexpected demand. Column (1d) reports the difference between surprise and expected demand multipliers at each level in column (1), showing that all differences are statistically significant at the 1% level.

Columns (4) through (6) estimate the results by sub-sample and find broadly similar results: price multipliers of predictable demand are, even if non-zero, an order of magnitude smaller than that associated with surprise demand. Apart from the last subsample where some standard errors are much wider, in all subsamples and for all levels, the multiplier of surprise demand is statistically significantly larger than that associated with expected demand. In unreported results, we also verify that the conclusion is not affected by using in-sample, rather than out-of-sample, predictions in conducting demand decomposition. Overall, consistent with theory, price impact primarily arises from the surprise component of demand.

Table F.14. Price impact of expected and surprise demand

We decompose weekly demand into expected and surprise components and estimate their price multipliers separately. Specifically, for demand in each level (3×3, 6×6, and idiosyncratic), we use 8 lags of demand and return to forecast demand using expanding windows, with the first quarter of the sample (1993 to 1999) as the initial "burn-in period". Using the specification similar to that in Table 3, the first three columns estimate price multipliers using the full sample while the last three columns use subsamples. Column (1d) reports the difference of multipliers between surprise and expected demand by level. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Dependent variable: weekly stock return ($\text{Ret}_{i,t}$)								
	Full sample					Sub-samples		
	Coefficient estimates				Difference	2000-2007	2008-2015	2016-2022
	(1)	(2)	(3)		(1d)	(4)	(5)	(6)
$\text{Demand}_{i,t}^{\text{expected, }3\times3}$	-1.351	-1.670		3×3	8.820***	-3.253	-0.018	3.288
	(1.859)	(2.341)			(1.961)	(2.821)	(1.824)	(4.295)
$\text{Demand}_{i,t}^{\text{expected, }6\times6}$	-1.761	-1.335		6×6	5.055***	-3.482	-1.001	1.892
	(1.211)	(1.335)			(1.232)	(2.140)	(0.981)	(1.579)
$\text{Demand}_{i,t}^{\text{expected, Idio}}$	0.098**	0.077*		Idio	1.473***	0.218***	-0.006	0.011
	(0.042)	(0.042)			(0.067)	(0.051)	(0.047)	(0.134)
$\text{Demand}_{i,t}^{\text{surprise, }3\times3}$	7.469***		7.473***			9.944***	5.094***	5.520***
	(0.626)		(0.629)			(0.949)	(0.534)	(1.750)
$\text{Demand}_{i,t}^{\text{surprise, }6\times6}$	3.293***		3.286***			4.388***	2.034***	2.353***
	(0.227)		(0.228)			(0.394)	(0.179)	(0.379)
$\text{Demand}_{i,t}^{\text{surprise, Idio}}$	1.571***		1.571***			2.103***	0.871***	1.233***
	(0.052)		(0.052)			(0.076)	(0.057)	(0.084)
Week FE	Y	Y	Y			Y	Y	Y
Other controls	Y	Y	Y			Y	Y	Y
Obs	2,921,729	2,921,729	2,921,729			1,169,628	977,086	775,015

F.2 Reconciling multiplier estimates by frequency

In tables 3 and 4, we estimate price multipliers at weekly, daily, and intraday frequencies. We can also estimate multipliers using slower data at monthly and quarterly frequencies. The resulting *prima facie* multiplier estimates are plotted by the red bars in Figure F.11 with 95% confidence intervals. While the results are similar across daily to weekly frequencies, the intraday multipliers are substantially higher. This has two possible explanations. It could be that intraday price pressures partially revert, or it could indicate stronger autocorrelation in intraday demand. In this section, we show support for the latter explanation.

As shown in Figure A.3, weekly demand exhibits little reversal or continuation. However, demand might exhibit stronger continuation at higher frequencies, in which case we should adjust the prima facie multiplier estimates $\tilde{M}$. Specifically, suppose each 1% of hourly demand is on average followed by another $\beta - 1$ units — so there is a total of β units of demand. Then, the adjusted hourly multiplier should be $M = \tilde{M}/\beta$. A similar adjustment is applied in G.5.2 of the May 2022 version of Gabaix and Koijen (2022) when examining price multipliers estimated using persistent microstructure order flow measures.

We now estimate the coefficient β at daily and intraday frequencies. For instance, at the daily frequency, we estimate scaling factors β at each level of aggregation (lev):

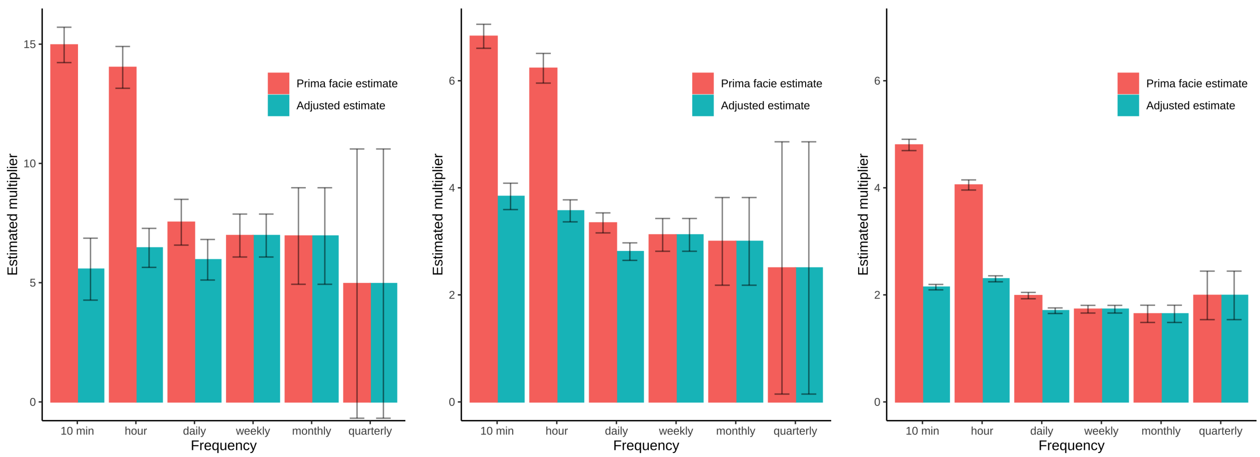
$$\sum_{h=0}^{4} \text{Demand}_{i,t+h}^{\text{lev}} = \beta^{\text{lev}} \cdot \text{Demand}_{i,t}^{\text{lev}} + \sum_{t=1}^{T} \eta_t^{\text{lev}} + \epsilon_{i,t}^{\text{lev}}$$

where $\{\eta_t^{\text{lev}}\}_t$ are time fixed effects and the independent variables are cumulative demand by the end of five trading days (a week). The resulting point estimates are $(\beta^{3\times3}, \beta^{6\times6}, \beta^{\text{Idio}}) = (1.26, 1.19, 1.17)$, indicating that there is slight persistence at the daily frequency, consistent with the findings in Table A.1. We also estimate the same for intraday frequencies by regressing cumulative one-week demand on high-frequency demand. The estimated scaling factors are higher at (2.17,1.75,1.76) and (2.69,1.78, 2.24) at hourly and 10-minute frequencies, respectively.

Figure F.11. Price multipliers by frequency

This figure plots price multipliers estimated by regressions with frequencies ranging from 10-minute to quarterly. The red bars represent prima facie estimates from regressing returns on contemporaneous demand components (e.g. Tables 3 and 4). The blue bars represent the adjusted estimates that account for the persistence of demand. The error bars represent 95% confidence intervals. The adjustment process is described in Appendix F.2.

(a) 3×3 style (b) 6×6 style (c) Idiosyncratic



We now adjust the multipliers for daily and higher frequencies ($M = \frac{\tilde{M}}{\beta}$) and plot them in the blue bars of Figure F.11. The standard errors are computed using the delta method that also takes into account estimation errors in β. We do not adjust estimates at slower frequencies as demand does not exhibit clear autocorrelation at those levels. As shown in the Figure, once the persistence of demand fluctuations is accounted for, the price multiplier estimates are similar across different observational frequencies.