In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis

Xavier Gabaix and Ralph S.J. Koijen*

February 4, 2021

Abstract

We develop a framework to theoretically and empirically analyze the fluctuations of the aggregate stock market. Households allocate capital to institutions, which are fairly constrained, for example operating with a mandate to maintain a fixed equity share or with moderate scope for variation. As a result, the price elasticity of demand of the aggregate stock market is small, so flows in and out of the stock market have large impacts on prices.

Using the recent method of granular instrumental variables, we find that investing $1 in the stock market increases the market's aggregate value by about $5. We also show that we can trace back the time variation in the market's volatility to flows and demand shocks of different investors.

We also analyze how key parts of macro-finance change if markets are inelastic. We show how general equilibrium models and pricing kernels can be generalized to incorporate flows, which makes them amenable to use in more realistic macroeconomic models, and to policy analysis. Our calibration implies that government purchases of equities have a non-trivial impact on prices. Corporate actions that would be neutral in a rational model, such as share buybacks, have substantial impacts too.

Our framework allows us to give a dynamic economic structure to old and recent datasets comprising holdings and flows in various segments of the market. The mystery of apparently random movements of the stock market, hard to link to fundamentals, is replaced by the more manageable problem of understanding the determinants of flows in inelastic markets. We delineate a research agenda that can explore a number of questions raised by this analysis, and might lead to a more concrete understanding of the origins of financial fluctuations across markets.

*xgabaix@fas.harvard.edu, ralph.koijen@chicagobooth.edu. We thank Ehsan Azarmsa, Antonio Coppola, Dong Ryeol Lee, Hae-Kang Lee, Simon Oh and Lingxuan Wu for excellent research assistance. We thank Francesca Bastianello, Jean-Philippe Bouchaud, Michael Brandt, John Campbell, Robin Greenwood, Lars Hansen, Sam Hanson, John Heaton, Tim Johnson, Arvind Krishnamurthy, Spencer Kwon, John Leahy, Hanno Lustig, Alan Moreira, Knut Mork, Toby Moskowitz, Stefan Nagel, Jonathan Parker, Lasse Pedersen, Joel Peress, Ivan Shaliastovich, Andrei Shleifer, Jeremy Stein, Larry Summers, Adi Sunderam, Jean Tirole, Harald Uhlig, Dimitri Vayanos, Motohiro Yogo, and participants at various seminars for comments. Gabaix thanks the Sloan Foundation for financial support. Koijen acknowledges financial support from the Center for Research in Security Prices at the University of Chicago Booth School of Business.

1 Introduction

One key open question is why the stock market exhibits so much volatility. This paper provides a new model and new evidence to suggest that this is because of demand flows in surprisingly inelastic markets. We make the case for this theoretically and empirically, and delineate some of the numerous implications of that perspective.

We start by asking a simple question: when an investor sells $1 worth of bonds, and buys $1 worth of stocks, what happens to the valuation of the aggregate stock market? In the simplest "efficient markets" model, the price is the present value of future dividends, so the valuation of the aggregate market should not change. However, we find both theoretically and empirically, using an instrumental variables strategy, that the market's aggregate value goes up by about $5 (for simplicity, we will use this simple round multiplier of 5 in the theory and discussion).[1] Hence, the stock market in this simple model is a very reactive economic machine, which turns an additional $1 of investment into an increase of $5 in aggregate market valuations.

Put another way, if investors create a flow of 1% as a fraction of the value of equities, the model implies that the value of the equity market goes up by 5%. This is the mirror image of the low aggregate price-elasticity of demand for stocks: if the price of the equity market portfolio goes up by 5%, demand falls by only 1%, so that the price elasticity is 0.2. In contrast, most rational or behavioral models would predict a very small impact, about 100 times smaller, and a price elasticity about 100 times larger.

This high sensitivity of prices to flows has large consequences: flows in the market affect market prices and expected returns in a quantitatively important way. We refer to this notion as the "inelastic markets hypothesis."

We lay out a simple model explaining market inelasticity. In its most basic version, a representative consumer can invest in two funds: a pure bond fund, and a mixed fund that invests in stocks and bonds according to a given mandate — for instance, that 80% of the fund's assets should be invested in equities. Then, we trace out what happens if the consumer sells $1 of the pure bond fund and invests this $1 in the mixed fund. The mixed fund must invest this inflow into stocks and bonds: but that pushes up the prices of stocks, which again makes the mixed fund want to invest more in stocks, which pushes prices up, and so on. In equilibrium, we find that the total value of the equity market increases by $5.

Then, the paper explores inelasticity in richer setups and finds that the ramifications of the simple model are robust. For instance, the core economics survives, suitably modified, if the fund is more actively contrarian, so that its policy is to buy more equities when the expected excess return on equities is high. It also holds in an infinite-horizon model: the price today is influenced by the cumulative inflows to date and the present value of future expected flows — divided again by the market elasticity. Moreover, the model aggregates well. If different investors have different elasticities, the total market elasticity is the size-weighted elasticity of market participants, size being the share of equity they hold. The model also clarifies how to measure flows into the aggregate stock market, which guides the empirical analysis.

The empirical core of this paper is to provide a quantification of the market's aggregate elasticity. To do that, we use a new instrumental variables approach, which was conceived for this paper and worked out in a stand-alone paper (Gabaix and Koijen (2020)), the "granular instrumental variables"

[1] This is linear and also works for sales: selling $2 worth of equities (buying $2 worth of bonds) decreases the valuation of aggregate equities by $10.

(GIV) approach. The key idea is that we use the idiosyncratic demand shocks of large institutions or sectors as a source of exogenous variation. We extract these idiosyncratic shocks from factor models estimated on the changes in holdings of various institutions and sectors. We then take the size-weighted sum of these idiosyncratic shocks (the GIV), and use it as a primitive instrument to see how these demand shocks affect aggregate prices and the demand of other investors. This way, we can estimate both the aggregate sensitivity of equity prices to demand shocks (which is the multiplier of around 5 we mentioned above) and the demand elasticity of various institutions (around 0.2).

Importantly, the data are consistent with a quite long-lasting price impact of flows. Indeed, in the simplest version of the model, the price impact is perfectly long-lasting. This is not because flows release information, but instead simply because the permanent shift in the demand for stocks must create a permanent shift in their equilibrium price. We perform a large number of robustness checks, for example using different data sets (the Flow of Funds as well as 13F filings). The findings are consistent across specifications, in the sense that the price impact multiplier remains around 5 — indeed, in the range of 3 to 8. Furthermore, lots of things fall into place then: for instance, the volatility of the market can be traced back to the volatility of flows and demand changes.

Here are three a priori reasons to entertain that markets would be inelastic First of all, if one wants to buy $1 worth of equities, many funds actually cannot supply that: for instance, a fund that invests entirely in equities cannot exchange them for bonds. Many institutions have tight mandates, something that we confirm in this paper. Relatedly, it is hard to find the "would-be arbitrageurs." For instance, hedge funds are small (they hold only about 5% of the equity market), and they tend to reduce their equity allocations in bad times (for example, because of outflows or risk constraints; see Ben-David et al. (2012)). Second, the transfer of equity risk across investor sectors is small (about 0.6% of the aggregate value of the equity market per quarter for the average pair of investor sectors). This suggests again that the demand elasticity of most investors is quite small. Third, a substantial literature has shown that when a company is removed from an index, its share price falls, where the latest estimates find a "micro" demand elasticity of approximately one (we give complete references in the literature review below). As the macro elasticity should arguably be lower than the micro elasticity (considering that, for example, Ford and General Motors are closer substitutes than the stock market index and a bond), this suggests a low macro elasticity, perhaps less than one.[2] Hence, even though our macro multiplier estimates are large, they are less surprising when compared to the evidence on micro elasticity estimates. Consistent with this argument, and in support of the inelastic markets hypothesis, Deuskar and Johnson (2011a) use high-frequency order flow data for S&P 500 futures to show that about half of the price variation can be attributed to flows shocks. Moreover, they find these shocks to be permanent over the

[2]Also, the empirical market microstructure estimates of price impact are larger than what we find: the price impact that the microstructure literature finds is a factor of about 15 (Frazzini et al. (2018)), which may make our estimate of 5 seem moderate. Microstructure results are typically couched in a form such as "buying 2% of the daily volume of a stock creates creates a permanent price increase of 0.12%". These estimates may seem to imply a small price impact, but work out to a price impact multiplier of $M = 15$. However, the interpretation is delicate, as we discuss in Section E.11. In short, this microstructure estimate of 15 has the interpretation, in inelastic markets with a micro elasticity of 1, that a large market-wide desired trade ("metaorder") is on average split into 15 smaller trades executed over time, by one or several institutions collectively (for example, by three funds pursuing a similar strategy, each splitting their desired position change into five smaller trades). Those microstructure results are also to be taken with caution, because identification is more difficult as trades are not exogenous. Using high frequency data with a GIV-based identification may be a promising way to enrich identification procedures in microstructure.

horizons that they consider.[3]

Suppose that the "inelastic markets hypothesis" is true; why do we care? First, flows are quantitatively impactful. We find that roughly over one third of all stock market fluctuations are driven by capital flows. As a result, one can replace the "dark matter" of asset pricing (whereby price movements are explained by hard-to-measure latent forces) with tangible flows and the demand shocks of different investors. We also link the time variation in the market's volatility to flows and demand shocks. This suggests a research program in which determinants of asset prices can be traced back to concrete flows by concrete investors. By studying the actions of these investors, we can infer their demand curves, and theorize about their determinants.

If equity markets are indeed inelastic, several questions that are irrelevant or uninteresting in traditional models become interesting. For instance, if the government buys stocks, stock prices go up — again by this factor of 5. This may be useful as a policy tool — a "quantitative easing" policy for stocks rather than long-term bonds. It may also be used to analyze previous policy experiments, in Hong Kong, Japan, and China, and give a quantitative framework to complement the previous qualitative discussions of policy proposals of this kind (Tobin (1998); Farmer (2010); Brunnermeier et al. (2020)).

Also, firms as financiers materially impact the market in our calibration. Prior research showed that firms react to price signals, such as in their decisions to issue dividends or raise funds in stocks versus bonds (Baker and Wurgler (2004); Ma (2019)): now we can quantity how firms' actions impact the market. For instance, stock buybacks can have a large aggregate effect. Suppose that the corporate sector buys back \$1 worth of equities rather than paying \$1 worth of dividends. In the traditional Modigliani-Miller world, the market value of equities does not change at all. In contrast, in an inelastic world, the value of equities goes up, by a tentative estimate of between \$0.5 and \$2.[4] As a naive non-economist might think, "if firms buy shares, that drives up the price of shares". A rational financial economist might say that this is sadly illiterate. But here the naive thinking is actually qualitatively correct in inelastic markets. Hence, potentially, as share buybacks account for a large portion of flows (they have been about as large as dividend payments in the recent decade), corporate actions account for a sizable share of equity purchases, and hence of the volatility and valuation growth of the stock market. This "corporate finance of inelastic markets" is an interesting avenue of research.

If markets are inelastic, then macro-finance should reflect that. Accordingly, we construct a general equilibrium model in the spirit of Lucas (1978) where there is a central role for flows and inelasticity. It clarifies the role of flows, the determination of the interest rate, and shows how to augment traditional general equilibrium models with flows in inelastic markets. That makes those models more realistic, and better suited for policy. This model may serve as a prototype for models enriched by inelasticity. Indeed, it calibrates well, and replicates quantitatively the salient features of the stock market, such as the volatility and size of the equity premium, the slow mean-reversion of the price-dividend ratio, and the ability to predict stock return with price-dividend ratio

[3]Deuskar and Johnson (2011a) study a system of equations in which flows may impact returns and returns may impact flows. To identify price impact, they rely on identification via heteroskedasticity as in Rigobon (2003). As only the demand shock in futures markets is used, and not in cash markets, we cannot directly translate the estimates into multipliers. However, under the assumption that flows in cash markets are highly correlated with flows in futures markets, it does show that flows explain a large fraction of market fluctuations, which is consistent with the inelastic markets hypothesis.

[4]The estimate is tentative, in part as it uses estimates of the rationality of the consumer after the buybacks.

at different horizons. We conclude that our general equilibrium model with "inelastic markets" is competitive with other widely-used general equilibrium models that match equity market moments, be it via habit formation (Campbell and Cochrane (1999)), long run risks (Bansal and Yaron (2004)), or variable rare disasters (Gabaix (2012), Wachter (2013)). In addition to proposing a new amplification mechanism, its main advantage, as we see it, is that is relies on an observable force, flows in and out of equities, this way diminishing the "dark matter" needed for asset pricing.

We also show how to marry flows to the "stochastic discount factor" (SDF) approach: the flows are primitive, and the SDF is a book-keeping device to record their influence on prices. This model could be helpful to get correct risk prices in macroeconomic models, including their variation due to flows.

One limitation of our study is that we postpone to future research the detailed investigation of what determines flows in the first place: instead, we provide descriptive statistics showing they correlate sensibly with other variables, such as news and measured beliefs. The reason is chiefly that this would be a stand-alone paper. But we think it is quite doable, and indeed we are working on this. Rather than studying "shocks to noise traders" abstractly, we replace them with investor-level flows and demand shocks that may be easier to understand: indeed, episode by episode, one can ask "why did firms lower their buybacks?" (answer: because they had lower earnings), "why did pension funds buy?" (answer: because their mandate forces them to buy stocks after stocks fall), or "why did hedge funds sell?" (answer: their investors sold, given their low past returns).

Our findings are not only at odds with traditional theoretical models, but also with the prevailing common wisdom in our profession at the time of writing this paper We quantify this via two surveys. We conducted a first survey by putting out a request via Twitter (using the #econtwitter tag) to complete an online survey. In addition, we asked participants of an online seminar at VirtualFinance.org to complete the same survey – this latter audience being naturally more representative of the population of academic researchers in finance. Both surveys were conducted before the paper was available online and before the seminar was conducted. We received 192 responses for the Twitter survey and 102 responses for the survey connected to the finance seminar.

The survey question was the following: "If a fund buys $1 billion worth of US equities (permanently; it sells bonds to finance that position), slowly over a quarter, how much does the aggregate market value of equities change?" The answer given in this paper is M times a billion, where M is the price impact multiplier, which we estimate to be around $M = 5$. At the same time, the simplest "efficient markets" answer is $M = 0$: in that view, the price is the present value of future dividends, unperturbed by flows.[5] In both surveys, the median answer was $M = 0$: surveyed economists, logically enough, rely on the basic finance model with efficient markets. The median positive answer was $M = 0.01$.[6] Hence, surveyed economists' priors expect the price impact of flows to deviate by a factor of over 500 from our calibrated and estimated values.

Literature review This paper is about the macro-elasticity of the market (that is, how the aggregate stock market's valuation increases if one buys $1 worth of stock by selling $1 worth of bonds). This is in contrast with the very large literature that studies the micro-elasticity of the

[5] In fact, the rational M is a bit higher than that; see Section 3.5.

[6] The answer $M \geq 1$ was given by only 2.5% of respondents in the Twitter survey and by 4% of respondents in the VirtualFinance.org survey. Section D.8 provides further details.

market (which describes how much the relative price of two stocks changes if one buys \$1 of one, and sells \$1 of the other): this literature includes Shleifer (1986), Harris and Gurel (1986a), Wurgler and Zhuravskaya (2002), Duffie (2010), Chang et al. (2014), and Koijen and Yogo (2019). A recent estimate by Chang et al. (2014) using index reconstitutions implies a micro elasticity of 1.5 using all benchmarked assets, and 0.4 when only using passive assets. Barbon and Gianinazzi (2019) estimate a longer-run demand elasticity using the quantitative easing program in Japan that involved equity purchases and estimate it to be equal to 1 as well. In contrast, our macro elasticity is about 5 times smaller. Samuelson (1998) opined that the stock market is not "macro-efficient" (it does not get the absolute value of the stock market as a whole correctly), though it might be "micro-efficient" (it gets the relative pricing of Ford versus General Motors roughly right). This paper is about the macro-efficiency of the market, while most of the above literature is about its micro-efficiency. The macro-elasticity of the market is a key parameter of interest, as we show in this paper, but there has been no attempt to estimate it.

A recent literature provides estimates of macro multipliers outside of the US using exogenous variation in demand due to institutional changes. Da et al. (2018) find a multiplier of 2.2 in Chile and Li et al. (2020a) find macro multipliers between 3.8 and 5.3 in China. In the US, Ben-David et al. (2020a) study a demand shock for style factors and estimate a multiplier of 5.3. As Ben-David et al. (2020a) rightly point out, it is plausible to expect the multiplier for style factors to fall in between the micro and the macro multiplier. Taken together, this evidence points to macro multipliers well above one, consistent with the inelastic markets hypothesis.

We share with Koijen and Yogo (2019) and Koijen et al. (2019) our reliance on holdings data by institutions, and the desire to estimate a demand function. We are mostly interested in the equilibrium in the aggregate stock market, as opposed to the cross-sectional focus of Koijen and Yogo (2019), and we emphasize the role of flows, and the dynamics of prices and capital flows over time. Using a similar modeling strategy as in Koijen and Yogo (2019), estimate a global demand system across global equity and bond markets to understand exchange rates, bond prices, and equity prices across countries.

We build on the insights of De Long et al. (1990), who write an equilibrium model in which noisy beliefs create demand shocks that move the market and the equity premium. De Long et al. (1990) discuss a rich set of qualitative ideas, some of which we can formally analyze and quantify, such as the failure of the Modigliani-Miller theorem and the notion that if most market participants passively hold the market portfolio, prices react sharply to flows. De Long et al. (1990) dealt with these issues qualitatively, but, influenced by it, a literature has studied the impact of mutual fund flows in the market, including Warther (1995) and Frazzini and Lamont (2008). One innovation of our paper is to provide a systematic quantitative framework to think about this, to include all sectors (not just mutual funds), and think about causal inference, via the GIV.

A few papers have modeled how flows might be important, examining general flows in currencies (Gabaix and Maggiori (2015), Greenwood et al. (2019), Gourinchas et al. (2020)), slow rebalancing mechanisms in currencies (Bacchetta and Van Wincoop (2010)) and equities (Chien et al. (2012)), or switching between types of stocks (Barberis and Shleifer (2003), Vayanos and Woolley (2013b)). However, we believe we are the first to conceptually and quantitatively explore the elasticity of the aggregate stock market using a simple economic model to link data on total holdings and flows to fluctuations in the aggregate stock market. We also provide the first instrumental variables estimate of the elasticity of the aggregate market. Camanho et al. (2019) provide a partial-equilibrium model of exchange rates with flows, quantified with the GIV methodology developed for the present paper and spelled out in Gabaix and Koijen (2020).

A related literature finds convincing evidence that supply and demand changes do affect prices and premia in partially segmented markets, for bonds (for example as in Greenwood and Vayanos (2014), Greenwood and Hanson (2013), and Vayanos and Vila (2020)), mortgage-backed securities (Gabaix et al. (2007)), or options (Garleanu et al. (2009)), with models which typically feature CARA investors and partial equilibrium. Here our focus is on stocks, while our model is quite different from the models in that literature (in particular, avoids CARA investors) and is also developed in general equilibrium. Our work also relates back to the work on flows and asset demand systems by Brainard and Tobin (1968) and Friedman (1977), among others. This literature faced two important challenges that we address; first, data on asset holdings were not as readily available as they are now and, second, there were no obvious methods to identify the slopes of asset demand curves. We also relate to the literature on slow-moving capital (Mitchell et al. (2007); Duffie (2010); Duffie and Strulovici (2012); Moreira (2019); Li (2018)), providing a new model for price impact with long-lasting effects, and an identified estimation. Finally, part of our contribution is a new model of intermediaries (He and Krishnamurthy (2013)), with a central role for flows, trading mandates, and inelasticity.

Much more distant to our paper is the theoretical microstructure literature (Kyle (1985)). There, inflows cause price changes, but crucially those inflows do not change the equity premium on average (as the mechanism is rational Bayesian updating, rather than limited risk-bearing capacity, unlike Kondor and Vayanos (2019)), and hence do not create excess volatility. In contrast, in our paper, inflows do change the equity premium, creating excessively volatile prices.

Outline Section 2 gives some simple suggestive facts showing who moves during market turmoils, and the size of flows. Section 3 develops our basic model of the stock market: it lays out the basic notions, and defines clearly elasticity and its link with price impact. It also gives the theoretical framework that we take to the data. Section 4 is the empirical analysis, including with an instrumental variable estimation of the aggregate market elasticity. Section 5 provides a general equilibrium model that helps to think about how everything fits together: it ties up the loose ends of the basic model of Section 3, and in particular endogenizes the interest rate and links cash flows to production and consumption. Section 6 gives a number of complements, for example on policy implications, corporate finance with inelastic markets, and the very long run determinants of flows. Section 7 provides a conclusion and thoughts about the research directions suggested by the present approach. The appendix contains the basic proofs, and details. The online appendix contains a number of robustness checks and extensions.

Notations We use $\mathcal{E}$ for equities, $\mathbb{E}$ for expectations, and E for equal-weighted averages. We call δ the average dividend-price ratio of the equity market. We generally use lowercase notations for deviations from a baseline. For a vector $X = (X_i)_{i=1\dots N}$ and a series of relative shares S_i with $\sum_{i=1}^N S_i = 1$, we let

$$X_E \coloneqq \frac{1}{N}\sum_{i=1}^N X_i, \qquad X_S \coloneqq \sum_{i=1}^N S_i X_i, \qquad X_\Gamma \coloneqq X_S - X_E, \tag{1}$$

so that X_E is the equal-weighted average of the vector's elements, X_S is the size-weighted average, and X_Γ is their difference. We define the mean of X_i (with $i = 1 \dots N$) with weights ω_i as:

$$\mathbb{E}_\omega [X_i] \coloneqq \frac{\sum_i \omega_i X_i}{\sum_i \omega_i}. \tag{2}$$

2 Initial Motivations to Explore the Inelastic Markets Hypothesis

In this section, we provide some suggestive facts that will motivate the inelastic market hypothesis, and the model. These facts are meant to be no more than suggestive: the core empirical results are in Section 4, in which we try to carefully quantify the key parameters of our model.

First, we document that institutions often have rigid equity shares in Section 2.1, and relatedly we seek to identify investors with elastic demand for the aggregate stock market in Section 2.2. That is, we ask: who are the deep-pocketed arbitrageurs that could make the aggregate stock market elastic? This question relates to the work by Brunnermeier and Nagel (2004), who show that hedge funds did not provide elasticity to the market during the technology bubble in the late nineties. Next, we show that the flows between sectors are relatively small and that small changes in quantities are correlated with large changes in prices, a sign of inelastic markets. We conclude by connecting the inelastic markets hypothesis to the evidence on micro elasticities, which also suggests that this hypothesis may hold. We describe the data that we use in Appendix C, which in this section is primarily from the Flow of Funds (FoF).

2.1 Institutions often have a quite rigid equity share

As a point of reference, we summarize in Figure 1 the evolution of ownership of the US equity market from 1993 (orange bars) to 2018 (green bars). During the last 25 years, equity ownership moved from households' direct holdings to institutions. The figure understates this trend as the "household sector" in the FoF includes various institutional investors such as hedge funds and non-profits (e.g., endowments). Broker dealers, who received much attention in the recent asset pricing literature, hold only a small fraction of the US equity market. This limits their ability to provide elasticity to the market.

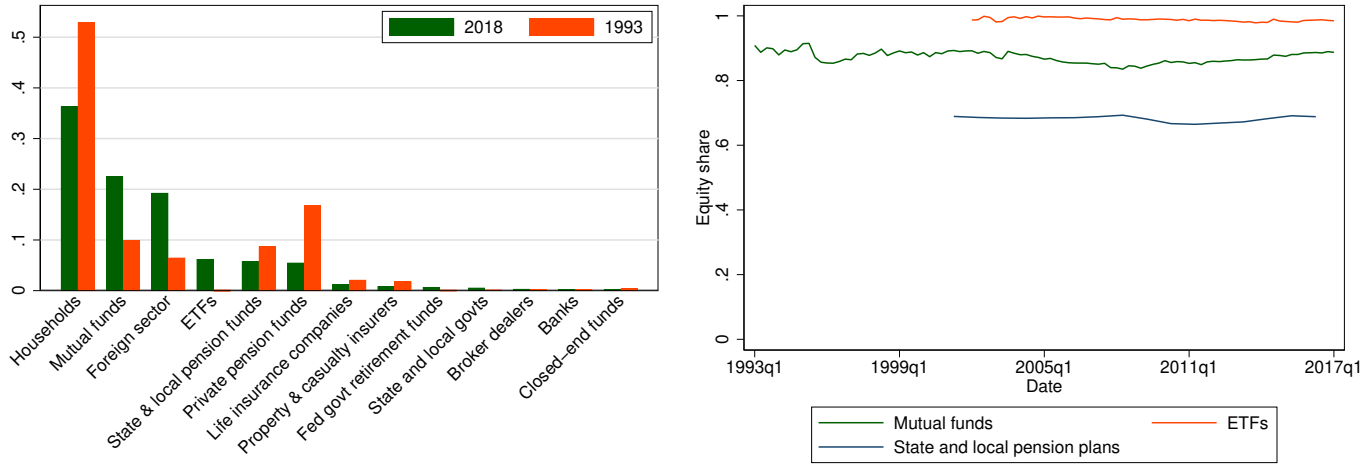
For some of these sectors, such as mutual funds, exchange-traded funds, and pension funds, we have investor-level data on equities and fixed income holdings,. In the right panel of Figure 1, we plot the equity share. The surprise is that equity shares are impressively stable over time.[7] This is consistent with many institutions being constrained to maintain a stable equity share, something that will motivate our model.

2.2 Flows between investor groups are small

If market fluctuations are the result of small demand or flow shocks hitting macro inelastic markets, studying extreme episodes may provide a hint as to which investor sectors have volatile demand and flow shocks and which investor sectors provide elasticity to the market. We therefore consider a case study of the two largest equity downturns in our sample, namely from 2000.Q2 to 2002.Q3 (the technology crash) and from 2007.Q4 to 2009.Q1 (the 2008 global financial crisis), as shown in Appendix D.2. To measure equity flows, we scale the dollar equity flows for each sector j, $\Delta F_{jt}^{\mathcal{E}}$, by the size of the aggregate market in the previous quarter, $\mathcal{E}_{t-1}$, $\frac{\Delta F_{jt}^{\mathcal{E}}}{\mathcal{E}_{t-1}}$. Importantly, we remove

[7] Consistent with the theory that we develop in the next section, we aggregate different investors in a given sector using the relative sizes of their equity portfolios (as opposed to assets under management). In particular for pension funds, they also invest in other asset classes, which lowers the average equity share. Throughout this paper, we focus on equity versus fixed income as the relevant margin of substitution, leaving other asset classes to future work.

Figure 1: Equity shares. The left panel of the figure plots the equity share in 1993 (orange bars) and in 2018 (green bars) by institutional sector. The right panel displays the value-weighted average equity share of mutual funds, ETFs, and state and local pension plans. The equity share of the different institutions are averaged using the relative equity size of each investor. The construction of the data is discussed in Appendix C.



the "mechanical" effects via revaluation, so that we show the "active" flows. We then average the percent flow by sector across quarters for a given downturn.

The left panel of Figure 2 corresponds to the tech crash and the right panel to the 2008 financial crisis. In the case of the 2008 financial crisis, we separately report the results for 2008.Q4, which is the worst quarterly return in our sample. In both cases, we select the eight sectors with the largest absolute flows as well as the corporate sector. While the total equity risk reallocation, on average per quarter, remains small, households sell about 0.5% of the market per quarter.[8]

During the 2008 financial crisis, net repurchases by firms fell (as firms cute their share buybacks in bad times) and indeed turned negative, implying that they issued equity. If we zoom in on 2008.Q4, we see large issuances (for instance by financial firms, in part forced by the government to issue shares),[9] which may have further amplified the market decline if the market is inelastic.

Who is providing elasticity to the market during these episodes? Quite surprisingly, the foreign sector as well as state and local pension funds are the sectors purchasing the most during each of the episodes. For the pension funds, thislikely reflects their mandate to maintain a fixed-share strategy instead of a conscious effort to time the market (see Proposition 2).[10]

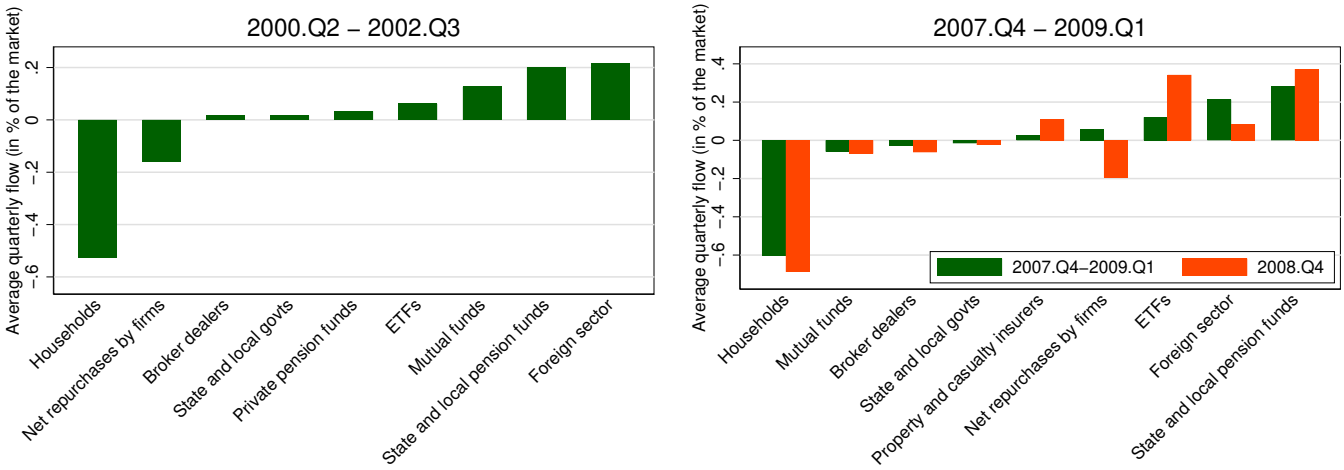
The flows across sectors are not only small during downturns, but also on average: total gross flows are only 0.85% per quarter (see Section D.3).

[8]We emphasize once more that the household sector in the FoF includes institutional investors such as hedge funds and non-profits (e.g., endowments), as it is computed as a residual.

[9]During this period, several firms received support from the government. In the FoF, new sectors were created that otherwise hold no equity positions. We adjust net repurchases and flows for these sectors in order not to distort our calculations; see Appendix C for details.

[10]Relatedly, Timmer (2018) finds that in German data, banks (broker dealers) sell when stock prices fall, and pension funds buy.

Figure 2: The figure illustrates the rebalancing of investors during drawdowns of the US stock market from 1993.Q1 to 2018.Q4. The left panel summarizes the data from the tech crash (from 2000.Q2 to 2002.Q3) and the right panel from the 2008 global financial crisis (from 2007.Q4 to 2009.Q1). We plot the average quarterly rebalancing by sector expressed as a fraction of the total market capitalization (expressed in %). In the right panel, we also replicate the calculation for the fourth quarter of 2008, which is the most negative quarterly return in our sample. In all cases, we select the eight sectors with the largest absolute flows as well as the corporate sector.



Hedge funds, as a whole, are small, and reduce equity holdings in downturns. We conclude by zooming in on one particular group of intermediaries that one can view as the prime example of smart arbitrageurs, namely hedge funds.[11] There are two reasons why hedge funds help little to increase the market's elasticity.

First, hedge funds own a small fraction of all US equities: they hold around 5% of the US stock market in 2016. Second, hedge funds reduce their equity holdings during equity downturns, as shown by Ben-David et al. (2012), which is the opposite of what we would expect to find if hedge funds are particularly macro price-elastic. Economically, their trading behavior may simply reflect that hedge fund investors withdraw their capital or that risk constraints bind,[12] which forces hedge funds to reduce their leverage and the overall riskiness of their portfolios. Still, the combination of their own risk constraints and the behavior of their clients leads hedge funds to be small, and to reduce equity holdings in downturns. Hence, hedge funds do not provide appreciable elasticity to markets in downturns.

[11] While a large literature explores the micro elasticity of hedge funds, we are interested in their market elasticity. In the FoF, hedge funds are part of the household sector and we cannot study them separately using these data. We can make progress using returns and 13F data, as we show below.

[12] See Adrian and Shin (2013) and Coimbra and Rey (2020) for models of intermediaries featuring value-at-risk constraints. In Appendix D.4 we use returns to estimate how hedge funds' betas move in response to the equity premium and we find that they reduce their equity share in bad times.

2.3 The macro elasticity should be lower than the micro elasticity, which is itself low

There is one more argument to think that the aggregate stock market is likely to be inelastic, namely the evidence on the micro elasticity of demand and the implications of theoretical models. We refer to the micro elasticity as the price effects resulting from exchanging one stock for another stock. The macro elasticity refers to the price movement due to moving capital from the bond market, say, to the stock market.[13] In most asset pricing theories, the micro elasticity exceeds the macro elasticity. Smart-money arbitrageurs such as hedge funds are eager to arbitrage idiosyncratic risks but are more reluctant to arbitrage the aggregate stock market if they think it is mispriced.

With this in mind, it is useful to connect to the literature on micro elasticities and in particular the index inclusion literature started by Harris and Gurel (1986a) and Shleifer (1986). Chang et al. (2014), for instance, find a demand elasticity around one using Russell index inclusions. Barbon and Gianinazzi (2019) find a similar demand elasticity of 1 in the longer-run using the quantitative easing program that involved equity purchases in Japan. This is much less elastic than what is implied by standard models for the micro elasticity, but in fact also much less elastic than what they imply for the macro elasticity, as we show in Section 3.5. However, if the theoretical prediction that the macro elasticity is below the micro elasticity does hold, then it seems plausible that macro markets are inelastic.

3 Basic Theory with Inelastic Markets

The traditional models of finance predict a macro elasticity that is considerably too high – something we shall soon spell out. Hence, we provide a model that we think is more realistic to think concretely about the determinants of stock demand. It is highly stylized, but will be useful to think about the determinants of elasticity (both conceptually and in terms of calibration) and to guide empirical work. We start with a two-period version, and then proceed to an infinite-horizon variant and a calibration.

3.1 Two-period model for the aggregate stock market

There is a representative stock in fixed supply of $Q^{\mathcal{E}}$ shares, and one riskless asset with a constant risk-free rate. The economy lasts for two periods $t = 0, 1$. The dividend D is paid at time 1. A representative consumer invests into stocks and bonds via two representative funds: a pure bond fund that only holds the riskless asset, and a mixed fund that holds both the riskless asset and the representative stock.[14]

At time 0, the mixed fund's assets under management (or equivalently wealth) are denoted by W, while P is the price of the stock, and Q^D is the mixed fund's demand for stocks expressed in quantity of shares. Therefore the fraction of the mixed fund's wealth invested in equities is $\frac{PQ^D}{W}$.

We consider that the mixed fund's demand for stocks is given by a mandate, saying that the it

[13] Section E.4 extends our model to formally define and analyze the micro elasticity.

[14] Those representative funds act competitively, i.e. are price takers.

should have a fraction invested in equities equal to:[15]

$$\frac{PQ^D}{W} = \theta e^{\kappa\hat{\pi}}, \tag{3}$$

where $\theta \in (0,1)$ is the typical equity share in the fund, while the rest is in the riskless bond. Calling $\pi = \frac{D^e}{P} - 1 - r_f$ the equity premium (with $D^e \coloneqq \mathbb{E}[D]$ the expected dividend at time 0), $\bar{\pi}$ is the average equity premium, and $\hat{\pi} \coloneqq \pi - \bar{\pi}$ is the deviation of the equity premium from its average.

In the simplest case, $\kappa = 0$, the fund has a fixed mandate to invest a fraction θ of its wealth in equities. When κ is strictly positive, the fund allocates more in equities when they have higher expected excess returns (hence, κ indexes how contrarian or forward-looking the fund is). This demand function appears sensible, and could be micro-founded along many lines – but to go straight to the effects we are interested in, we take it as an exogenous mandate.[16]

If consumers were fully rational, the mandate would not matter: consumers could undo all mechanical impacts of the mandate. But consumers will not be fully rational, so mandates will have an impact.

Prices react a lot to flows in inelastic markets We use bars to denote the value at time $t = 0^-$, before any shocks. At that time 0^-, the mixed fund has wealth $\bar{W}$, and holds all the equity shares. We assume that before the shocks, equities have an equity premium $\bar{\pi}$, so that the dividend-price ratio is at its corresponding value, $\delta = \frac{\bar{D}^e}{\bar{P}}$, where $\bar{P}$, $\bar{D}^e$ are the baseline values for the stock's price and the expected dividend.

Then, at time 0, we suppose that the representative consumer decides (for whatever reason) to sell ΔF dollars of the pure bond fund, and invest them in the mixed fund. What happens then? The answer is given in the next proposition. [17]

Proposition 1. *Suppose that the representative consumer sells ΔF dollars of the pure bond fund, and invests them in the mixed fund, so that the inflow in the mixed fund is a fraction $f = \frac{\Delta F}{\bar{W}}$ of the fund's value. Then, the stock price changes by a fraction $p \coloneqq \frac{P-\bar{P}}{\bar{P}}$ equal to:*

$$p = \frac{f}{\zeta}, \tag{4}$$

where ζ is the macro elasticity of demand:

$$\zeta = 1 - \theta + \kappa\delta. \tag{5}$$

This illustrates that flows can have large price impacts if the price elasticity of demand ζ is sufficiently low, and shows the key role of this price elasticity, which is the center of this paper.

We shall soon calibrate $\zeta \simeq 0.2$, so that $\frac{1}{\zeta} \simeq 5$. Then, (4) means that if investors buy 1% of the equity market (selling bonds), the price of equities increases by 5%. This is symmetric and

[15] We write the mandate in "number of shares", but it is equivalent to a "fraction of assets invested in equity" formulation.

[16] This representative mixed fund's mandate can be viewed as a stand-in for other frictions such as inertia, or a rule of thumb that a behavioral household might follow for its stock allocation. As a result, the institutionalization of the market does not necessarily result in more inelasticity; indeed, the opposite will be true if households managing their own portfolios are more inert than institutions.

[17] It is exact for $\kappa = 0$, and uses a first-order Taylor expansion for small flows f when $\kappa \neq 0$. Following common practice in macro-finance, we do Taylor expansions of the leading terms, omitting the formal mentions of $O(\cdot)$ terms.

linear: if investors sell 2% of the market, the price of equities falls by 10%. One can also say it in dollar terms. If someone buys \$1 worth of equities (selling \$1 worth of bonds), the market value of aggregate equities increases by \$5. This great ability of the market to multiply shocks will play a central role, and indeed will be a source of the "excess volatility" of stocks.

To see where (4) comes from, we derive the demand for stocks. We do that under a slightly more general assumption, assuming that at time 0 there may be a change d in the value of fundamentals. We call q^D and d the fractional deviations of the equity demand and expected dividend from their baseline values:

$$q^D = \frac{Q^D}{\bar{Q}^D} - 1, \qquad d = \frac{D^e}{\bar{D}^e} - 1. \tag{6}$$

We perform the analysis for small disturbances f, d, and hence small p, q^D, here and throughout the paper.

Proposition 2. (Demand for aggregate equities in the two-period model) *The demand change (compared to the baseline) is*

$$q^D = -\zeta p + f + \kappa \delta d, \tag{7}$$

where ζ is the macro elasticity of demand given in (5)*.*

Proposition 2 delivers Proposition 1 as a simple consequence. Indeed, as the supply of shares does not change, we must have $q^D = 0$ in the equilibrium after the flow shock. This and (7), with $d = 0$, give the price, $p = \frac{f}{\zeta}$.

Proof of Proposition 2 Before the inflow shock, the mixed fund's wealth is $\bar{W} = \bar{P}\bar{Q}^D + \bar{B}$, where $\bar{P}\bar{Q}^D$ and $\bar{B}$ are respectively the fund's holdings of equities and bonds:

$$\bar{P}\bar{Q}^D = \theta \bar{W}, \qquad \bar{B} = (1-\theta)\,\bar{W}.$$

After the inflow shock, the fund's wealth is $W = P\bar{Q}^D + \bar{B} + \Delta F$ as the price has changed to P, and the fund's wealth is made of the equity value $P\bar{Q}^D$ (at the new price P) and the inflow ΔF, so that

$$\Delta W = (\Delta P)\,\bar{Q}^D + \Delta F. \tag{8}$$

So, the value of the assets in the fund changes by a fraction:

$$w \coloneqq \frac{\Delta W}{\bar{W}} = \frac{(\Delta P)\,\bar{Q}^D}{\bar{W}} + \frac{\Delta F}{\bar{W}} = \frac{\bar{P}\bar{Q}^D}{\bar{W}} \times \frac{(\Delta P)}{\bar{P}} + f = \theta \times p + f,$$

that is

$$w = \theta p + f. \tag{9}$$

This means that the value of the fund increases via the inflow of f, and via the appreciation of the stock p, to which the fund has an exposure θ.

Let us first take the case $\kappa = 0$. The demand (3) is

$$Q^D = \frac{\theta W}{P} = \frac{\theta \bar{W}\,(1+w)}{\bar{P}\,(1+p)} = \bar{Q}^D \frac{1+w}{1+p},$$

so that the fractional change in the fund's demand for shares is:

$$q^D = \frac{Q^D}{\bar{Q}^D} - 1 = \frac{w-p}{1+p} = \frac{\theta p + f - p}{1+p} = \frac{f - \zeta p}{1+p},$$

with $\zeta = 1 - \theta$. We see how ζ is the Marshallian demand elasticity, which includes crucial income effects. For small price changes, this gives $q^D \simeq f - \zeta p$. We also see that, when $\kappa = 0$, the equilibrium condition $q^D = 0$ leads to $p = \frac{f}{\zeta}$ exactly. Finally, the case $\kappa > 0$ is in Appendix A. □

We now comment on the economics of Proposition 1.

An undergraduate example To think through the economics of Proposition 1, we found the following simple, undergraduate-level example useful. Suppose that the representative mixed fund always holds 80% in equities ($\theta = 0.8, \kappa = 0$), so that $\zeta = 0.2$ and $\frac{1}{\zeta} = 5$. Then an extra 1% inflow into the stock market increases the total market valuation by 5%.

It is instructive to think through the logic of this example. Suppose that the representative mixed fund starts with \$80 in stocks (of which there are 80 shares, worth \$1 each) and \$20 in bonds. There are also $\$B$ worth of bonds outstanding. Suppose now that an outside investor sells \$1 of bonds from the pure bond fund (he had $\$B - \$20$ in the pure bond fund, and now he has $\$B - \$21$), and invests this \$1 into the mixed fund. In terms of "direct impact", there is a \$0.8 extra demand for the stock (equal to 1% of the stock market valuation), and \$0.2 for the bonds. But that is before market equilibrium forces kick in.

What is the final outcome? In equilibrium, the pure bond fund still holds $\$B - \$21$ worth of bonds. The balanced fund's holdings are \$21 in bonds (indeed, it holds the remaining \$21 of bonds) and $4 \times \$21 = \$84$ in stocks (as the balanced fund keeps a 4:1 ratio of stocks to bonds, the value of the stocks it holds must be \$84). As the balanced fund holds all 80 shares, the stock price is $P = \frac{\$84}{80} = \$1.05$, whereas it started at $P = \$1$: stock prices have increased by 5%. The fund's value also has increased by 5%, to \$105.

We see that the increase in stock prices is indeed by a factor $\frac{1}{\zeta} = \frac{1}{1-\theta} = 5$. Only \$0.8 was invested in equities, yet the value of the equity market increased by \$4, again a five-fold multiplier.[18]

Demand shocks for equities not mediated by the fund have the same price impact $\frac{1}{\zeta}$ Suppose that there is a flow f^D into stocks, in the sense that a segment of the market inelastically buys a fraction f^D of the shares. Then, this purchase of f^D percent of the shares increases the price of equities by $\frac{f^D}{\zeta}$ percent.[19]

Prices are partially myopic to future fundamentals in inelastic markets We complete the picture by examining what happens when at time 0 there is a change in the expected dividend by a fraction d.[20]

Proposition 3. (Price equilibrium) *After a flow f and a change in the expected dividend d, the share price moves by a fractional amount equal to*

[18] Anticipating implementation issues, we note that if there is no change in supply, a basic question arises on how to measure flows. In this example, we can only identify the flow into markets by looking at the flow into the outside asset that is assumed to be perfectly elastic. Based on the equity market itself, we only notice the price movement, but this could be due to changes in expected fundamentals or other demand shocks.

[19] To see this, we suppose that there is no flow into the mixed fund nor news about dividends ($f = d = 0$). As in equilibrium the total demand change must be zero, the equilibrium is $f^D + q^D = 0$ with $q^D = -\zeta p$, by equation (7). That yields $p = \frac{f^D}{\zeta}$.

[20] The proof is very simple. Market clearing implies that total demand does not change, $q^D = 0$. This and (7) gives the price change, $p = \frac{f}{\zeta} + \frac{\kappa\delta}{\zeta} d$.

$$p = \frac{f}{\zeta} + M^D d, \tag{10}$$

where M^D is the market's effective attention to future fundamentals:

$$M^D = \frac{\kappa\delta}{\zeta} = \frac{\kappa\delta}{1-\theta+\kappa\delta} \in [0,1]\,. \tag{11}$$

Indeed, if there is a shock to expected fundamentals, $d = \frac{\Delta D^e}{\bar{D}^e}$, then by (10) the equity price moves less than one-for-one with expected fundamentals, as long as there is no flow f from the households correcting this:

$$p = M^D d.$$

This partial myopia is a second and more minor force in inelastic markets, and is the flip side of inelasticity. The market is very forward-looking ($M^D \to 1$) only when $\kappa \to \infty$ (so that investors buy a lot of stocks when the equity premium is a bit higher than usual), or $\theta \to 1$, and $M^D = 0$ when the market doesn't react to fundamentals (which happens for example if traders do not react to the equity premium, $\kappa \to 0$).

In this example, the expected dividend causes a price change "without flows". Some price changes are "caused" by aggregate flows f, but those caused by news about future disturbances (here d) move prices without aggregate flows.

3.2 Infinite horizon model for the aggregate stock market

We extend the static model to a dynamic one. The forces will generalize, in an empirically implementable way. There is again a constant risk-free rate r_f, taken here to be exogenous (Section 5 endogenizes it in general equilibrium, but here we concentrate on the core economics of inelasticity). The representative stock gives a dividend D_t.

We assume a slightly more general mandate: it says that the fraction invested in equities, $\frac{P_t Q_t^D}{W_t}$, should be

$$\frac{P_t Q_t^D}{W_t} = \theta e^{\kappa\hat{\pi}_t + \nu_t}, \tag{12}$$

where as before $\hat{\pi}_t \coloneqq \pi_t - \bar{\pi}$ is the deviation of the equity premium from its average, and we allow for additional demand shocks, ν_t. These can be thought of as taste shocks, or as a proxy for information.

We define by $\bar{P}_t$, $\bar{D}_t, \bar{W}_t$, and $\bar{Q}_t^D$ the baseline price, dividend, wealth, and quantity of shares held by the mixed fund – the values before any flow or extra news friction.[21] We call p_t, w_t, d_t, q_t^D the deviations from the baseline, so that $d_t = \frac{D_t}{\bar{D}_t} - 1$, $p_t = \frac{P_t}{\bar{P}_t} - 1$, $w_t = \frac{W_t}{\bar{W}_t} - 1$, and $q_t^D = \frac{Q_t^D}{\bar{Q}_t^D} - 1$. We define the flow f_t as the *sum* of past fractional inflows, since a date $t = 0$ when the market was in equilibrium ($\hat{\pi}_0 = 0$):

$$f_t = \sum_{s=0}^{t} \frac{\Delta F_s}{\bar{W}_{s-1}}, \tag{13}$$

[21] For simplicity, we postpone the more formal definition of those "baseline" values, but they are exactly defined in the context of the economy of Proposition 7 and its proof.

where ΔF_s are dollar inflows. In this definition, we assume that dividends are passed to consumers, which implies that reinvested dividends count as new flows. We call again δ the typical dividend-price ratio, $\delta \coloneqq \overline{D/P} = r_f + \bar{\pi} - g$, with g the growth rate of dividends. We call the expected dividend deviation $d_t^e = \mathbb{E}_t d_{t+1}$. The expected excess return is $\pi_t = \frac{\mathbb{E}_t[\Delta P_{t+1} + D_{t+1}]}{P_t} - r$, and we use the Taylor expansion (see Section E.1 for a derivation)

$$\hat{\pi}_t = \delta \left(d_t^e - p_t \right) + \mathbb{E}_t \left[\Delta p_{t+1} \right]. \tag{14}$$

The aggregate demand for stocks is as follows, generalizing (7).

Proposition 4. (Demand for aggregate equities in the infinite-horizon model) *The demand change for equities (compared to the baseline) is*

$$q_t^D = -\zeta p_t + f_t + \nu_t + \kappa \delta d_t^e + \kappa \mathbb{E}_t \left[\Delta p_{t+1} \right], \tag{15}$$

where $\zeta = 1 - \theta + \kappa\delta$ is the aggregate elasticity of the demand for stocks, as in (5).

If there is no change in supply, the equilibrium condition is given by $q_t^D = 0$. This yields the stock price as follows (the proof is in Appendix A).

Proposition 5. (Equilibrium price in the infinite-horizon model) *The equilibrium price of aggregate equities is (expressed as a deviation from the baseline):*

$$p_t = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho}{(1+\rho)^{\tau-t+1}} \left(\frac{f_\tau + \nu_\tau}{\zeta} + M^D d_\tau^e \right) \tag{16}$$

where $\rho = \frac{\zeta}{\kappa}$ is the "macro market effective discount rate",

$$\rho = \frac{\zeta}{\kappa} = \delta + \frac{1-\theta}{\kappa}, \tag{17}$$

and $M^D = \frac{\kappa\delta}{1-\theta+\kappa\delta} \in [0,1]$ is a "coefficient of partial reaction to future fundamentals." The deviation of the equity premium from its average is:

$$\hat{\pi}_t = \frac{(1-\theta)\, p_t - (f_t + \nu_t)}{\kappa}. \tag{18}$$

We next analyze the economics of Proposition 5. The classical (or undergraduate) "efficient markets" benchmark, where the risk premium is kept constant by very strong arbitrage forces, corresponds to $\kappa = \infty$, so that $\zeta = \infty$ and $\rho = \delta$.

In (16), the price discounts future dividends at a rate $\rho \geq \delta$ given in (17).[22] So, *the market is more myopic (higher ρ) when it is less sensitive to the equity premium (lower κ) and when the mixed fund has a lower equity share (lower θ).*[23] It makes good sense that a lower sensitivity to the equity premium makes the market less reactive to the future, hence more myopic.[24,25] In the rest of this section, we set $\nu_t = 0$; the general case simply comes from replacing f_t by $f_t + \nu_t$.

[22] This myopia in (16) generates momentum: because the market is myopic (by (16)), dividend news are only slowly incorporated into the price.

[23] The formula extends to changes in the interest rate, as in $r_{ft} = \bar{r}_f + \hat{r}_{ft}$. As (14) becomes $\hat{\pi}_t = \mathbb{E}_t \Delta p_{t+1} + \delta \left(d_t^e - p_t \right) - \hat{r}_{ft}$, all expressions are the same, replacing d_t^e by $d_t^e - \frac{1}{\delta}\hat{r}_{ft}$, including in (16).

[24] The intuition for the sign of the impact of θ on ρ is as follow: The extra term $\frac{1-\theta}{\kappa}$ in $\rho = \delta + \frac{1-\theta}{\kappa}$ is the ratio of the "present looking" (myopic) demand elasticity $1-\theta$ to the "forward looking" elasticity κ. Hence a higher θ leads to a less myopic demand.

[25] Here the demand (12) depends on the equity premium as $\kappa\hat{\pi}_t = \kappa \mathbb{E}_t \Delta p_{t+1} + \kappa\delta \left(d_t^e - p_t \right)$. A variant would be

A permanent inflow has a permanent effect on the price and future expected returns of equities Suppose that at time 0 there is an inflow f_0 that does not mean-revert. Then, the impact on the time-t price is (via (16), with $\mathbb{E}_0[f_\tau] = f_0$):

$$\mathbb{E}_0[p_t] = \frac{1}{\zeta} f_0. \tag{19}$$

So, the "price impact" is permanent. As a result, the equity premium is permanently lower, $\hat{\pi} = -\delta \frac{f_0}{\zeta}$. Again, this is not due to any informational channel.[26] This is simply because, if the equity demand has permanently increased, equity prices should be permanently higher.

Quantitatively, if prices increase by 10% due to uninformed flows, the per annum expected excess return falls by a mere 0.3% (indeed, assuming a dividend yield of 3%, $\hat{\pi} = -\delta p = -(3\%) \times 0.1 = 0.3\%$) This is a vivid reminder that the absence of detectable market timing strategies tells us little about market efficiency (Shiller (1984)). Similarly, Black (1986) famously argued that the aggregate stock market can be mispriced by as much as a factor of two; in our model, if this is due to a permanent inflow, that would lead to a 2% change in the expected excess return,[27] which is less than a single standard error deviation of the expected excess return estimate if one were to use 30 years of data.

The impact of a mean-reverting flow Suppose now that at time 0 there is an inflow f_0 that mean-reverts at a rate $\phi_f \in [0,1]$, so that the cumulative flow is $\mathbb{E}_0[f_\tau] = (1-\phi_f)^\tau f_0$. Then, if there are no further disturbances, the impact on the time-t price is $p_t = \frac{1}{\zeta + \kappa\phi_f} f_t$ (see (16)), implying

$$\mathbb{E}_0[p_t] = \frac{(1-\phi_f)^t}{\zeta + \kappa\phi_f} f_0, \tag{20}$$

and the change in the equity premium is $\hat{\pi}_t = -\frac{\delta+\phi}{\zeta+\kappa\phi} f_t$ (see (18)). Hence, an inflow that has faster mean reversion leads to a smaller change in the price of equities (compared to a permanent inflow), but a *larger* change in their equity premium (indeed, $\frac{\delta+\phi_f}{\zeta+\kappa\phi_f}$ is increasing in ϕ_f). Those effects dissipate as the inflow mean-reverts, at a rate ϕ_f.

Predictable future inflows or changes in fundamentals create predictable price drifts Suppose that it is announced at time 0 that a permanent inflow f_T will happen at time $T > 0$. The price impact for $t \in [0,T]$ is $p_t = \frac{1}{(1+\rho)^T} \frac{f_T}{\zeta}$ (see (16), using $f_\tau = 1_{\tau \geq T} f_T$), so that after the initial jump, the price gradually drifts upward (assuming that the inflow is positive, for concreteness). Hence, the risk premium is elevated by $\hat{\pi}_t = \frac{1-\theta}{\kappa} p_t$ (for $t \in [0,T)$, see (18)), and more elevated as one nears the inflow. After the inflow, though, we are back to the case of a permanently elevated price and permanent lower equity premium ($p_t = \frac{f_T}{\zeta}$ and $\hat{\pi}_t = -\delta \frac{f_T}{\zeta}$ for $t \geq T$). The same price drift before the shock happens for a predictable increase in future fundamentals such as dividends.

---

that investors "see" the price dividend ratio as differently predictive from the expected price movement, so that in their demand we equalize $\kappa\hat{\pi}_t$ with $\kappa\mathbb{E}_t\Delta p_{t+1} + \kappa^D \delta (d_t^e - p_t)$ where potentially $\kappa^D \neq \kappa$ (e.g., if "tangible" predictors are deemed more reliable, $\kappa < \kappa^D$). Then the demand elasticity is $\zeta = 1 - \theta + \kappa^D\delta$, the effective discount factor is $\rho = \frac{\zeta}{\kappa}$, and (16) still holds, with $M^D = \frac{\kappa^D\delta}{1-\theta+\kappa^D\delta}$. This highlights that κ^D increases market elasticity ζ, while κ increases market "forward-lookingness" $\frac{1}{\rho}$.

[26] In a Kyle (1985) model, flows change prices, like in our model; but they do not on change the equity premium (on average), which is a crucial difference with our model. Section E.11 details the link with the Kyle model.

[27] Indeed, with $p = \ln 2$, $\hat{\pi} = -\delta p = -(3\%) \times 0.7 \simeq -2\%$.

A simple benchmark To think about the stochastic steady state, the following benchmark is useful. Cumulative past flows are decomposed as:

$$f_t = (1-\theta)\, d_t + \tilde{f}_t, \tag{21}$$

where d_t is the past cumulative dividend growth, taken to be a martingale, and $\tilde{f}_t$ is autoregressive with speed of mean-reversion ϕ_f, $\tilde{f}_t = (1-\phi_f)\,\tilde{f}_{t-1} + \varepsilon_t^f$ with $\mathbb{E}_{t-1}\left[\varepsilon_t^f\right] = 0$. Then, (16) gives:

$$p_t = b_f^p \tilde{f}_t + d_t, \qquad \hat{\pi}_t = b_f^\pi \tilde{f}_t, \qquad b_f^p = \frac{1}{\zeta + \kappa\phi_f}, \qquad b_f^\pi = -\left(\delta + \phi_f\right) b_f^p. \tag{22}$$

A high inflow increases equity prices and hence lowers the equity premium.

3.3 Heterogeneous funds

This subsection will show that with heterogeneous funds, everything (such as the equilibrium price) goes through as above. We simply need to replace quantities like θ, κ, f by their averages over funds, weighing those funds by the dollar value of their equity holdings. We now detail this.

Two-period model There are I funds (I as in Institutions) indexed by i. Suppose funds hold just the stock market and the short-term bond. An inflow ΔF_i goes to fund i and we let $f_i = \frac{\Delta F_i}{W_i}$ so that the fund's demand change is as in (7):

$$q_i^D = -\zeta_i p + \kappa_i \delta d + f_i,$$

with $\zeta_i = 1 - \theta_i + \kappa_i\delta$. Total demand for stocks is $Q = \sum_i Q_i \left(1 + q_i^D\right)$. We call $\mathcal{E}_i$ the equity holdings (in dollars) of fund i, $\mathcal{E}_i = Q_i P = \theta_i W_i$, and $S_i = \frac{\mathcal{E}_i}{\sum_j \mathcal{E}_j}$ is the share of total equities held by fund i. Finally, for a variable x_i ($i = 1 \ldots I$), we define $x_S \coloneqq \sum_i S_i x_i$, the size-weighted mean of x_i. So, the aggregate demand change is:

$$q_S^D = -\zeta_S p + \kappa_S \delta d + f_S.$$

So, we have the same expressions as in the basic model (which had $q^D = -\zeta p + f + \kappa\delta d$, see (7)), replacing θ, κ, f_t, and ν_t by their the equity-weighted averages:

$$\theta = \theta_S, \qquad \kappa = \kappa_S, \qquad \zeta = \zeta_S, \qquad f = f_S. \tag{23}$$

For instance, the price change after a size-weighted flow $f_S = \sum_i S_i f_i$ is (when there is no dividend news, so $d = 0$) $p = \frac{f_S}{\zeta_S}$, as in (4).

The equity-weighted equity share always exceeds the wealth-weighted equity share, $\theta_S \geq \mathbb{E}_W[\theta_i]$.[28] The former is what matters for the macro elasticity, while the latter is directly available in aggregated data. This makes the disaggregation issues potentially non-trivial, and will require some care in the empirical part.

[28] Indeed, using $\mathcal{E}_i = \theta_i W_i$, $\theta_S = \mathbb{E}_S[\theta_i] = \sum_i S_i\theta_i = \frac{\sum_i \mathcal{E}_i\theta_i}{\sum_i \mathcal{E}_i} = \frac{\sum_i W_i\theta_i^2}{\sum_i W_i\theta_i} = \frac{\mathbb{E}_W\left[\theta_i^2\right]}{\mathbb{E}_W[\theta_i]} \geq \mathbb{E}_W[\theta_i]$.

Infinite-horizon model Fund i has demand deviation (15), $q_{it}^D = -\zeta_i p_t + f_{it} + \nu_{it} + \kappa_i \delta d_t^e + \kappa_i \mathbb{E}_t [\Delta p_{t+1}]$. Aggregating over all funds i, we have (taking again the equity-weighted averages):

$$q_{St}^D = -\zeta_S p_t + f_{St} + \nu_{St} + \kappa_S \delta d_t^e + \kappa_S \mathbb{E}_t [\Delta p_{t+1}].$$

This means that (15) holds, replacing θ, κ, f_t, and ν_t by their the equity-weighted averages as in (23). As a result, the rest of the results in that section also hold (for example including (16)), using equity-share weighted averages.

Share repurchases and issuances are just a type of flow Suppose that corporations buy back shares, meaning that they buy:

$$f_C = \frac{\text{Net repurchases (in value)}}{\text{Total equity value}} = -\frac{\text{Net issuances (in value)}}{\text{Total equity value}}. \tag{24}$$

Then, in the basic two-period model (the same holds in the infinite-horizon model) the basic net demand for shares is as above, using the total flow:

$$f \coloneqq f_S + f_C, \tag{25}$$

which is equal to the size-weighted total flow in the funds, f_S, plus share repurchases (as a fraction of the market value of equities). In short, on top of the traditional flows of investors into equities, we want to add share repurchases by corporations. In addition, if firms have a supply elasticity ζ_C, then the basic equilibrium is: $f_S - \zeta p = -f^C + \zeta_C p$ (this is, change in demand $f_S - \zeta p$ equal change in supply $-f^C + \zeta_C p$), so that $p = \frac{f_S + f_C}{\zeta + \zeta_C}$, and the effective market elasticity is $\zeta + \zeta_C$.

3.4 A calibration

We provide a tentative calibration for the aggregate equity market. Relying on Section 3.3, we could imagine that only a fraction m_p of funds are "active", while the rest are simply buy-and-hold, and hence provide no elasticity to the market. Then, the basic formula becomes (by the analysis in Section 3.3):

$$\zeta = m_p (1 - \theta + \kappa_a \delta), \tag{26}$$

where κ_a is the forward-lookingness of active funds.[29] We want to understand the value of $\zeta \simeq 0.2$ that we already discussed.

Some parameters in (26) are easy to estimate in the micro data. We take a dividend-price ratio $\delta = \frac{D}{P} = 4\%/\text{year}$ (indeed, we use yearly units throughout).[30] Given the results in Figure 1, we take an equity share $\theta = 0.85$ (equity-weighted in the sense of Section 3.3).

We set $m_p = 0.95$, which seems conservative, as no inertia would mean $m_p = 1$.[31] This leads to imputing a value of $\kappa = \kappa_a m_p = 1.5$ years. The market discount rate (17) is then:

$$\rho = \frac{\zeta}{\kappa} = 13\%/\text{year} \tag{27}$$

[29] In Section E.12, we consider a dynamic model of inertia, which leads to a different short- and long-run elasticity. However, unless investors are very inert, the quantitative impact is small. We therefore take the limiting case here where a fraction $1 - m_p$ is buy-and-hold investors.

[30] Conceptually, our dividends are "total payouts", so are a bit higher than actual dividends. This is not very important quantitively for the value of ζ.

[31] We could without difficulty set $m_p = 1$, but having in mind future extensions, it seems useful to keep m_p as a parameter that can be calibrated or estimated.

so that the market's myopia is quite moderate (if the market were fully forward looking, we would have $\rho = \delta$).

We were led to a conjecture of $\kappa \simeq 1.5$ years, but estimating the risk-premium elasticity κ directly from the data is quite hard.[32] Here, we provide a few thought experiments and descriptive statistics showing that $\kappa \simeq 1.5$ is plausible given the micro evidence.

First, we perform a few thought experiments to see what we might expect κ to be. The simplest rational model of portfolio choice where $\theta_{it} = \frac{\pi_t}{\gamma_i \sigma^2}$ gives $\kappa = \frac{d\theta_{it}}{d\pi_t}\frac{1}{\theta_{it}} = \frac{1}{\pi_t} = 20$, using an annual equity premium of 5%.[33] But, as we shall see, we do not observe such large swings in investors' portfolios. The frictionless rational model predicts agents that are much too reactive, like in much of this paper, and in much of economics (Gabaix (2019)).

To get a further feel for κ, suppose the equity premium increases from $\pi = 5\%$ to $\pi = 10\%$, which is a shift equal to about one to two standard deviations of its unconditional time-series variation (Cochrane (2011); Martin (2017)). A very flexible fund with an average equity share of 50% might change its equity allocation from 50% to 75%. This flexible fund would have $\kappa_i = \frac{d\theta_i}{d\pi}\frac{1}{\theta_i} = \frac{0.25}{0.05}\frac{1}{0.5} = 10$. However, these are large swings in a fund's strategic asset allocation that are not typically observed empirically, so that they are at most valid only for very flexible investors. As many balanced funds have a fixed-share mandate and $\kappa = 0$, we hypothesize a κ for a typical fund with equity share of 50% equal to about 5. Moreover, a 100% equity funds needs to have $\kappa = 0$: more generally, the rigidity mechanically should increase with the equity share θ_i. So, we might tentatively parametrize a typical value of κ as $\kappa_i = K(1 - \theta_i)$, with $K \simeq 10$. So, we obtain $\kappa = \kappa_S = H(1 - \theta_S) = 10 \times 0.15 = 1.5$. This gives a simple microeconomic interpretation for the value $\kappa \simeq 1.5$.

Simple regressions also suggests a low κ. Figure 1 already showed a very high stability of equity shares. Appendix D.4 quantifies this more systematically for hedge funds, state and local pension funds, ETFs, and mutual funds. In all cases, we find a small κ, and indeed sometimes a negative κ, as generated by positive feedback trading. This evidence is just for a few important groups of investors, and is limited by the availability of detailed data on their equity and fixed income position, but we conjecture that our results plausibly extend to other classes of investors.

Hence, both the extant micro evidence pointing to fairly stable equity shares, and the macro evidence provided in this paper of a low macro elasticity, point to the view that κ (the sensitivity of the aggregate allocation to the equity premium) is fairly low. Together with the relatively stringent mandates (some of which could be behavioral – like a self-imposed desire not to actively rebalance), this gives rise to a very low market elasticity ζ.

When flows are mean-reverting, the price impact is $M = \frac{1}{\zeta^M}$, with $\zeta^M = \zeta + \kappa\phi_f$ (see (20) and (22)). We calibrate $\phi_f = 4\%$/yearto match the speed of mean-reversion of the dividend/price ratio.[34] That leads to $\zeta^M - \zeta = 6\%$, so that the difference is fairly minor. That leads to a price

[32]The GIV allows to infer the impact of changes in prices on quantities and vice-versa. But estimating κ separately from ζ means estimating how a change in the risk premia – controlling for the change in prices — affects positions. Given that most changes affect risk premia and prices simultaneously, this is a difficult task. It would be possible to do this, for example, by exploiting clear unexpected announcements about future dividends or flows.

[33]This is for a fund maximizing rationally a CRRA function of financial wealth. In Section 3.5 we consider a more sophisticated thought experiment, with a consumer maximizing lifetime utility out of labor income in additional to financial wealth. Then, the value of ζ is even higher.

[34]We compute the dividend yield by summing dividends during the last 12 months relative to the current level of the CRSP value-weighted return index from January 1947 to December 2018. The annual autocorrelation of the log dividend yield during this sample is equal to $\rho^{OLS} = 0.91$ with OLS standard errors equal to 0.048. We then remove the Kendall (1954) bias $\frac{4-3\rho}{T}$ over our sample of $T = 72$ years, which is around $\frac{4}{72}$. Thus we calibrate

impact $M = \frac{1}{\zeta^M} = 4$ when flows mean-revert.

3.5 Traditional rational or behavioral models predict that markets should be extremely price-elastic

In this section, we contrast our findings with the typical macro demand elasticities implied by most frictionless rational or behavioral models, and find that these are strongly inconsistent with the low price-elasticities that we model and estimate empirically.

First, as a partial intuition, if agents were risk neutral and the equity premium were 0, any price discrepancy would lead to an arbitrage, and the price elasticity of demand would be infinite, $\zeta^r = \infty$. This is the intuition behind the most basic form of the efficient markets hypothesis, where the price is always equal to the present value of dividends (with a constant discount rate), independently of flows.

Second, let us examine the more sophisticated case with risk-averse agents. We model aggregate income Y_t as going to the equity dividend as $D_t = \psi Y_t$, and the rest going to labor and other forms of business as $\Omega_t = (1-\psi) Y_t$. For simplicity, we consider the most classic case: the consumer has utility $\sum_t e^{-\rho t} \frac{C_t^{1-\gamma}}{1-\gamma}$ and the endowment Y_t has i.i.d. growth, $Y_t = G_t Y_{t-1}$. The basic case is the lognormal one, $G_t = e^{g\Delta t + \sigma \varepsilon_t - \frac{\sigma^2}{2}\Delta t}$ (with ε_t a standard Gaussian variable). We also consider a disaster model, where $G_t = e^{g\Delta t}$ if there is no disaster (which happens with probability $1 - p^D \Delta t$), and $G_t = e^{g\Delta t} B$ if there is a disaster (which happens with probability $p^D \Delta t$), so that if there is a disaster, the economy shrinks by a factor $B \in (0, 1]$.

Suppose that for some reason the market value of equities is different from its rational level, permanently, by a fraction p – that is, the price of equities is permanently $P_t = P_t^* (1+p)$, where P_t^* is the rational price.[35] How much capital should flow into equities? The next proposition answers this (the proof is in Section E.1).

Proposition 6. (Market elasticity in frictionless rational or behavioral models) *We derive the Marshallian price-elasticity of the demand for stocks in two classes of frictionless models. We suppose that all agents are frictionless (and with common beliefs, which can be rational or behavioral), with CRRA utility and with i.i.d. endowment growth. In the basic model in which growth rates are lognormal, the elasticity of demand for equities is:*

$$\zeta^r = \frac{1}{\pi} \frac{C}{W^{\mathcal{E}}}, \tag{28}$$

where π is the equity premium, C is aggregate consumption and $W^{\mathcal{E}}$ is the stock market capitalization. In a disaster model where growth rates follow a jump process, the elasticity of demand for

$\phi_f = 1 - \rho^{OLS} - \frac{4}{72} \simeq 4\%$.

[35] Johnson (2006) introduces a definition of market illiquidity that pertains to asset pricing models, whether or not there is trade between agents. His measure quantifies the equilibrium price change induced by a perturbation in asset supplies. Johnson (2006) examines this measure in the context of several rational setups, including a Lucas model, and this measure of illiquidity can be large and variable. We cannot use his results here because his definition of liquidity allows the interest rate to change when equity prices change, unlike our Marshallian demand elasticity ζ, which holds for a given interest rate. His notion of liquidity is generally lower than our elasticity, sometimes by an infinite factor. Indeed, take the case where there is no risk (the equity market is mispriced, so that the price is not the discounted value of the known dividend). In our model, the elasticity of demand is infinite (see Proposition 6), as it would be in most models with a riskless arbitrage opportunity, but Johnson's liquidity measure remains finite. In addition, we account for human capital, which is absent in Johnson's definition, and is quantitatively important.

equities is $\zeta^{r,D} = \frac{1}{\pi}\frac{C}{W^{\mathcal{E}}}\frac{(1-B^{\gamma})B}{\gamma(1-B)}$*, where* B *is the recovery rate of the endowment after a disaster.*

Take the calibrated values $C = 0.8Y$, where Y is GDP, $W^{\mathcal{E}} = Y$ (as the typical market capitalization is roughly equal to GDP), and $\pi = 4\%$. Then (28) implies that the elasticity predicted by rational models is $\zeta^r = 20$. Hence, with a calibrated and empirical elasticity $\zeta = 0.2$, we find that the basic rational model predicts an elasticity of demand 100 times bigger than the empirical one:

$$\frac{\zeta^r}{\zeta} = 100. \tag{29}$$

Summing up, we find that *frictionless rational or behavioral models (of the common "wrong beliefs" type) predict an elasticity of demand 100 times bigger than the calibrated and empirical one*. Indeed, in a behavioral model agents may have wrong beliefs, but they strongly act on their beliefs, with the same elasticity as in rational models (replacing the equity premium π in (28) by the perceived equity premium, but both are typically calibrated to have the same average value).

Now take the disaster model. Using the above calibration and the values $B = 2/3$, $\gamma = 4$ (Barro (2006), Gabaix (2012)), the elasticity in a disaster model given by Proposition 6 is $\zeta^{r,D} = 8$, so that it is 40 times larger than the empirical one. We strongly suspect that a similar reasoning would work for a habit formation model (Campbell and Cochrane (1999)) and a long run risks model (Bansal and Yaron (2004)).[36,37]

4 Estimating the Aggregate Market Elasticity

The previous sections illustrate the importance of estimating the elasticity of the aggregate stock market. Estimating this parameter is a challenge, as is the case for most elasticities in macroeconomics. In the context of asset pricing, large literatures try to estimate the coefficient of relative risk aversion, the elasticity of inter-temporal substitution, and the micro-elasticity of demand, but not the macro elasticity. In this section, we provide first estimates of the macro elasticity of the US stock market using the method developed in Gabaix and Koijen (2020), called Granular Instrumental Variables (GIV). However, given the relevance of this parameter, we believe it would be valuable for future empirical asset pricing research to explore different estimation and identification strategies in estimating its value.

In Section 4.1, we provide the basic intuition behind the GIV estimator. We report the estimates in Section 4.2 using sector-level data from the Flow of Funds and in Section 4.3 using investor-level data using 13F filings. In Section 4.4, we trace the volatility of aggregate stock market returns back to different categories of investors, and we highlight in particular the role of sector-specific shocks. We explore the connection between capital flows and asset prices in Section 4.5. We also connect capital flows to macro-economic variables and measures of beliefs to provide an initial analysis of the potential determinants of flows into the equity market.

[36] One could imagine other models, with idiosyncratic risk, but that would take us far afield.

[37] As a correlate, traditional models counterfactually predict very correlated flows and beliefs. Indeed, with several institutions and a demand $q_{it} = -\zeta^r p_t + f^{\nu}_{it}$, and $\zeta^r \simeq 20$, the term $\zeta^r p_t$ has an annual volatility of $\zeta^r \sigma_r = 20 \times 0.15 = 3$, or 300% per year. However, the annual volatility of equity holdings changes, we have seen, is about $\sigma_{q_i} \simeq 2\%$. Hence, to account for the empirical facts, we would need extremely volatile flows and demand changes f^{ν}_{it} of about 300%. In contrast, empirical flows f_{it} (when they can be measured) are about 1%. Hence, we would need almost perfectly correlated news and taste shocks ν_{it}, of 300% per year. All of this strikes us as quite implausible. It seems like a very difficult challenge to fit our facts with a traditional model.

Data sources We combine data from various sources that we briefly outline here; we leave a detailed description for Appendix C. The sample period used varies depending on the data source, and we discuss our choices in terms of data construction and sample selection in the appendix as well. In short, we combine sector-level holdings data from the US Flow of Funds (FoF), equity data from CRSP, data on firm fundamentals from Compustat, 13F investor-level holdings data from FactSet (as in Koijen et al. (2019)), detailed portfolio data for mutual funds and exchange-traded funds (ETFs) from Morningstar, fund-level portfolio and return data for state and local pension funds from The Center for Retirement Research at Boston College, hedge fund index returns data from Hedge Fund Research, Inc., macro data from the US Bureau of Economic Analysis, and survey data on return expectations from Gallup.

4.1 Intuition behind the GIV estimator

We first provide a brief summary of the GIV method – the appendix and Gabaix and Koijen (2020) provide many more details, such as a justification of efficiency and extensions. Recall that we use the following notations, with the shares S_j adding up to 1:

$$X_E \coloneqq \frac{1}{N}\sum_{j=1}^{N} X_j, \qquad X_S \coloneqq \sum_{j=1}^{N} S_j X_j, \qquad X_\Gamma \coloneqq X_S - X_E. \tag{30}$$

Suppose that we have a time series of changes in investors' equity holdings, Δq_{jt}, which can be modeled as (omitting constants):[38]

$$\Delta q_{jt} = -\zeta \Delta p_t + f^{\nu}_{jt}, \tag{31}$$

where Δp_t is the log price change of aggregate equities, and ζ is the demand elasticity of interest — we take it as constant across sectors in this introduction, but will relax this in Section 4.3. We consider the following model for f^{ν}_{jt}:

$$f^{\nu}_{jt} = \lambda_j' \eta_t + u_{jt}, \tag{32}$$

where η_t is a vector of common shocks (which can include observable factors, such as GDP growth, or latent factors), λ_j is a vector of factor loading, and u_{jt} is an idiosyncratic shock. We make throughout the key identification assumption that

$$\mathbb{E}\left[u_{jt}\eta_t\right] = 0. \tag{33}$$

The GIV method identifies ζ using variation that comes from the idiosyncratic shocks, u_{jt}.

Using market clearing,we have $\Delta q_{St} = 0$, that is

$$\Delta p_t = M\left(\lambda_S'\eta_t + u_{St}\right),$$

for the multiplier

$$M = \frac{1}{\zeta}.$$

The goal is to estimate M, which identifies the demand elasticity, ζ.

[38] To lighten things up, we simplify a bit the notations. Compared to (15), we use the notation f^{ν}_{jt} for $\Delta f^{\nu}_{jt} \coloneqq \Delta f_{jt} + \Delta\nu_{jt} + \kappa_j \Delta\mathbb{E}_t\left[\delta d^e_t + \Delta p_{t+1}\right]$. Later, we absorb the change-in-expectation terms $\kappa_j \Delta\mathbb{E}_t\left[\delta d^e_t + \Delta p_{t+1}\right]$ into the "demand shifter" $\Delta\nu_{jt}$.

The basic idea of the GIV is the following. We use idiosyncratic shocks to demand, u_{jt}, as primitive disturbances to the system, and we see how they affect prices and quantities. The GIV is the size-weighted sum of those idiosyncratic shocks. Indeed, if we had access to u_{St}, we could just estimate M by OLS, regressing $\Delta p_t = M u_{St} + \varepsilon_t$. We next detail how to obtain those idiosyncratic shocks, or at least good proxies for them that make the above reasoning valid.

Simple example with uniform loadings We start with the case where there is a single factor, η_t, and $\lambda_j = 1$, so that all loadings on the common shock are uniform. Then, the GIV is constructed from data as follows:

$$Z_t \coloneqq \Delta q_{\Gamma t} \coloneqq \Delta q_{St} - \Delta q_{Et}.$$

As $\Delta q_{St} = -\zeta \Delta p_t + \eta_t + u_{St}$ and $\Delta q_{Et} = -\zeta \Delta p_t + \eta_t + u_{Et}$, we have:

$$Z_t = u_{St} - u_{Et} =: u_{\Gamma t}. \tag{34}$$

Hence, $Z_t = u_{\Gamma t}$ is a combination of idiosyncratic shocks only, so that it is uncorrelated with η_t. This orthogonality condition makes $u_{\Gamma t}$ a valid instrument: it is our GIV. Furthermore, if u_{jt} is homoskedastic, then $u_{\Gamma t}$ is uncorrelated with u_{Et} (the same condition holds in the more general case of uncorrelated heteroskedastic u_{jt}, with the pseudo-equal weight, so $Z_t \coloneqq \Delta q_{St} - \Delta q_{\tilde{E}t}$). This implies that $\Delta p_t = M u_{\Gamma t} + e_t$, where $e_t = M(\eta_t + u_{Et})$ is uncorrelated with Z_t. Hence, if we estimate the OLS regression

$$\Delta p_t = M Z_t + e_t, \tag{35}$$

then this identifies the true multiplier M. Alternatively, we can estimate ζ directly using Z_t as an instrumental variable for Δp_t in the regression

$$\Delta q_{Et} = -\zeta \Delta p_t + \epsilon_t, \qquad \text{with } \Delta p_t \text{ instrumented by } Z_t \tag{36}$$

Intuitively, we use the sector-specific, or idiosyncratic, demand shocks of one sector as a source of exogenous price variation to estimate the demand elasticity of another sector.[39]

We emphasize that the methodology works even if we do not have data on flows f_{jt} – it is enough to have data on equity holdings q_{jt}. This implies that we identify idiosyncratic shocks to $f^{\nu}_{jt} = f_{jt} + \nu_{jt}$, where f_{jt} are capital flows and ν_{jt} are demand shocks.[40]

General case with non-uniform loadings In the general case with non-uniform loadings and an r-dimensional vector of common latent shocks η_t, we define $\check{a}_{jt} \coloneqq a_{jt} - a_{Et}$, that is, the cross-sectionally demeaned value of a vector a_t. We run a principal component analysis (PCA) via the model

$$\Delta \check{q}_{jt} = \check{\lambda}_j' \eta_t + \check{u}_{it}. \tag{37}$$

[39] Viewed this way, the GIV estimator generalizes the idea behind the index exclusion literature to estimate the micro elasticity. In the index inclusion literature, a demand shock to the group of index investors (assuming the inclusion of a stock into the index is random) can be used to estimate the slope of the demand curve of the non-index investors.

[40] If we were to have high-quality data on capital flows, f_{jt}, then we could construct another granular instrumental variable using capital flows by extracting idiosyncratic shocks to f_{jt}. However, our theory implies that we need accurate data on equity and bond holdings to measure capital flows correctly, which are unavailable in the US. Fortunately, however, we can implement the GIV procedure using Δq_{jt}, which does not require knowledge of holdings in other assets than equities.

In this way, we extract r principal components, η_t. Then, we run the following OLS regression, using the extracted factors η_t as controls:

$$\Delta p_t = MZ_t + \beta' \eta_t + e_t, \tag{38}$$

and read off the multiplier M as the coefficient on the GIV Z_t. The rest of Gabaix and Koijen (2020) discusses numerous extensions of this basic structure, and show how it is optimal by various metrics. As before, we can estimate ζ directly using Z_t as an instrumental variable for Δp_t in the regression

$$\Delta q_{Et} = -\zeta \Delta p_t + \beta' \eta_t + \epsilon_t.$$

We leave the technical details of the specific algorithms that we use to Appendix B.1. We demonstrate the performance of the GIV estimator in our specific setting in Appendix D.1 using simulations.

GIV: Requirements and threats to identification For the GIV to be consistent, we need (33) to hold: the idea is that there are random "bets" or "shocks" to various fund managers, institutions and sectors, that are orthogonal to all reasonable common macro factors such as GDP, TFP, and so forth. For the GIV to be a powerful instrument, we need large idiosyncratic shocks, and a few large institutions, so that the market is "granular" in the sense that the idiosyncratic trading shocks of a few large players meaningfully affect the aggregate.[41] Fortunately, this is verified in our setting, as it is in related settings in macro (Gabaix (2011), Carvalho and Grassi (2019)), trade (Di Giovanni and Levchenko (2012)) or finance (Amiti and Weinstein (2018), Herskovic et al. (forthcoming), Galaasen et al. (2020)).

The main threats to identification with GIV are that we do not properly control for common factors, or that the loadings on the omitted factor are correlated with size, such that $\lambda_S - \lambda_E \neq 0$. To mitigate the risk of omitted factors, we extract additional factors and explore the stability of the estimates as we add extra factors. We also perform an overidentification test.

When firms are elastic and flows mean-revert When firms have a supply elasticity ζ_C, the total elasticity is $\zeta + \zeta_C$, as we saw in Section 3.3. When flows mean-revert with speed ϕ_f, the measured elasticity is $\zeta + \kappa\phi_f$, as we saw in (20) and (22). Combining those two enrichments, the price measured impact

$$M = \frac{1}{\zeta^M}, \qquad \zeta^M \coloneqq \zeta + \kappa\phi_f + \zeta_C \tag{39}$$

As $\kappa\phi_f$ and ζ_C appear to be very small, the difference between ζ, $\zeta + \kappa\phi_f$ and ζ^M is rather minor, and is best kept out of mind in the first pass. Still, to be completely explicit, when we measure "ζ" we measure a quantity that is $\zeta + \kappa\phi_f$ if flows mean-revert at speed ϕ_f, and is strictly speaking ζ when they do not mean-revert. [42]

[41] Indeed, when flow shocks have volatility σ_u, $var(u_S) = H\sigma_u^2$, with $H = \sum_j S_j^2$. This "Herfindahl" H of the holdings shares must be high: so we need a few large entities, such as funds or sectors.

[42] Another enrichment would to make the flows sensitive to contemporaneous returns, say with an semi-elasticity ζ^f, as in $\Delta f_t = -\zeta^f \Delta p_t + \Delta \tilde{f}_t$. If $\zeta^q = 1 - \theta + \kappa\delta$ is the elasticity of the fund's holdings given flows, the elasticity of the funds' holdings is $\zeta = \zeta^q + \zeta^f$. If we have only holdings data (but not flow data), we can measure $\zeta^q + \zeta^f$. If we have flow data, we can measure also ζ^f directly, via the GIV.

Table 1: Estimates of the macro elasticity. The first two columns report estimates of M with one and two principal components, η_t, respectively. The next two columns report the elasticity estimates, ζ, regressing the equal-weighted change in equity holdings Δq_E on the price change Δp instrumented by the GIV Z The next two columns report the elasticity of the corporate sector, ζ_C. The final column reports the estimates of the same specification as the first column, but we omit Z_t, where $Z_t = \sum_j S_{j,t-1}\Delta\check{q}_{jt}$ and $\Delta\check{q}_{jt}$ defined in (77), to estimate the impact of sector-specific shocks on prices. In constructing $\Delta\check{q}_{jt}$, all estimates control for quarterly GDP growth. We report the standard errors, which are corrected for autocorrelation, in parentheses. The sample is from 1993.Q1 to 2018.Q4. The final line reports the R-squared values for the OLS regressions.

	Δp	Δp	Δq_E	Δq_E	Δq_C	Δq_C	Δp
Z	7.08	5.28					
	(1.86)	(1.10)					
Δp			-0.13	-0.17	-0.01	-0.01	
			(0.04)	(0.05)	(0.01)	(0.02)	
GDP growth	5.99	5.97	0.56	0.85	0.22	0.23	5.93
	(0.69)	(0.67)	(0.27)	(0.33)	(0.13)	(0.16)	(0.91)
η_1	21.06	23.72	3.98	5.49	-0.72	-0.64	31.50
	(13.58)	(12.79)	(2.08)	(2.07)	(0.69)	(0.81)	(15.57)
η_2		29.95		5.62		0.29	
		(6.54)		(2.15)		(0.67)	
Constant	-0.01	-0.01	0.00	0.00	-0.00	-0.00	-0.01
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Observations	104	104	104	104	104	104	104
R^2	0.436	0.515					0.279

4.2 Elasticity estimates using sector-level data

Benchmark estimates We first report the GIV estimates of the macro elasticity using data from 1993.Q1 to 2018.Q4 using the Flow of Funds (FoF). Throughout this section, we model investors' demand as

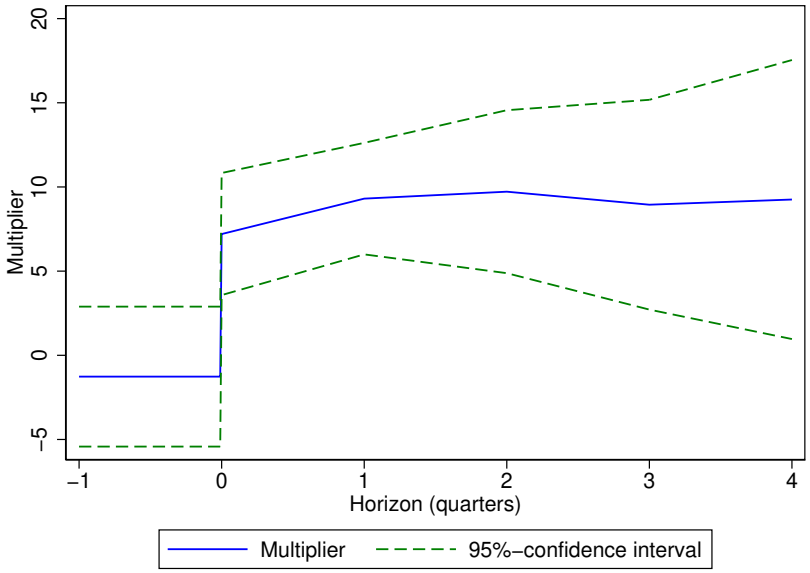
$$\Delta q_{jt} = \alpha_j - \zeta \Delta p_t + \lambda_j' \eta_t + u_{jt}, \tag{40}$$

where we assume that the demand elasticity is the same across sectors. We relax this assumption below using 13F data. We consider the same model for the corporate sector, but allow for a different demand elasticity, ζ_C. The vector η_t includes GDP growth, a time trend,[43] and one or more latent factors, η_t^{PC}.

The results are presented in Table 1. The first column reports the estimates where we use a single principal component to isolate the idiosyncratic shocks to various sectors, in addition to a common factor on which all sectors load equally.

[43]We include a time trend as some sectors grew faster in the nineties, for instance, than in the later period. We show the robustness of our results to not including the time trend.

Figure 3: Estimates of the aggregate multiplier $M = \frac{1}{\zeta}$ by horizon. The figure plots the multi-period impact of demand shocks: a demand shock of f_t at date t increases the (log) price of equities from $t-1$ to $t+h$ by Mf_t. We use the GIV for instrumentation, see (41). The horizontal axis indicates the horizon in quarters, from zero (that is, the current) to four quarters. Standard errors are adjusted for autocorrelation. The sample is from 1993.Q1 to 2018.Q4.



We estimate a multiplier of $M = 7.1$, implying that purchasing 1% of the market results in a 7.1% change in prices. The corresponding standard error is 1.9.[44] In the second column, we add a second principal component. This lowers the multiplier estimate to $M = 5.3$. That is, purchasing 1% of the market results in a 5.3% change in prices. Both estimates imply that the aggregate stock market is quite macro inelastic.

In the next two columns, we estimate the elasticities, ζ, by regressing demand changes on instrumented changes in prices, as in (36). With one principal component, we estimate an elasticity of $\zeta = 0.13$ and with two principal components, we estimate $\zeta = 0.17$. In the next two columns, we estimate the supply elasticity of the corporate sector. The short-run elasticity is low and equals 0.01 for both one and two principal components.[45] This implies that the combined elasticity is 0.14 (with one principal component) or 0.18 (with two principal components). The corresponding multipliers, $M = \frac{1}{\zeta+\zeta_C}$, are $M = 7.1$ and $M = 5.9$, respectively.

In the final column, we report the same regression as in the first column but without the instrument Z_t. By comparing the R-squared values, we obtain an estimate of the importance of sector-specific shocks on prices. We find that the difference in R-squared values is 16%, which highlights the importance of sector-specific shocks on prices.

[44] We report Newey-West standard errors using the bandwidth selection as in Newey and West (1994).

[45] This small contemporaneous elasticity of the supply of shares by the corporate sector, estimated here causally by IV, is consistent with the OLS findings of Ma (2019). She finds (Table VII) that $\frac{\text{Gross equity issuance}}{\text{Assets}} = 0.01\hat{\pi}$ (plus other terms) at the quarterly frequency. Using that equity is about two thirds of assets, this leads, at the annual frequency, to $\Delta q_C = \frac{3}{2} \cdot 4 \cdot 0.01\hat{\pi} = 0.06\hat{\pi}$, so that (by (14)) $\zeta_C = \delta \times 0.06 = 0.0024$. However, these estimates do not rule out the possibility that the medium- or long-run elasticities are higher and that firms play an important role in stabilizing asset prices.

The impact of flows at longer horizons In Figure 3, we explore how demand and flow shocks propagate across time. To this end, we extend the earlier analysis by estimating

$$p_{t+h} - p_{t-1} = a_h + M_h Z_t + c_h \eta_t^{PC,e} + d_h \Delta y_t + \epsilon_{t+h}, \tag{41}$$

for $h = 0, 1, \ldots, 4$ quarters. Recall that $\eta_t^{PC,e}$ is the principal component, extracted in the third step of the GIV procedure as outlined in Section B.1. The figure reports M_h at a certain horizon. We also consider a regression where we replace the left-hand side by $p_{t-1} - p_{t-2}$, which we refer to as $h = -1$. To construct the confidence intervals, we account for the autocorrelation in the residuals due to overlapping data.

We find that the cumulative impact is stable over time. This is intuitive as sharp reversals would imply a strong negative autocorrelation in returns, which is not something that we observe for the aggregate stock market at a quarterly frequency. As such, and consistent with the theory, persistent flow shocks lead to persistent deviations in prices. Size-weighted sector-specific demand shocks are also not correlated with returns at $t-1$ (that is, $h = -1$). Unfortunately, however, the confidence interval widens quite rapidly with the horizon, which limits what we can say about the long-run multiplier.

Robustness We explore the robustness of our estimates along various dimensions. In the interest of space, we report and discuss the tables in Appendix D.5. In Tables D.8–D.10, we consider a variety of robustness checks related to the sample period, data construction, and implementation choices of the GIV estimator. We conclude that our results are robust to these changes in the empirical strategy with multiplier estimates ranging from 3.5 to 8.

4.3 Elasticity estimates using investor-level data

We provide an alternative estimate of the same elasticity, but now using more disaggregated, investor-level 13F data.[46] We use 13F data from FactSet that covers the period from 1999.Q2 to 2017.Q2. The 13F data has two potential advantages over FoF data. First, we have a large cross-section of investors, which allows us to accurately extract latent factors. Second, we can construct investor-level characteristics, such as their assets under management (AUM), turnover, and characteristics of their investment strategy (e.g., with respect to size, value, and market beta). This allows us to estimate the model with heterogeneous demand elasticities across investors and uncover common factors associated with those characteristics (e.g., common flows from value- to growth-oriented, technology funds). The disadvantage, however, is that we do not have fixed income holdings and we cannot make the connection to capital flows. In addition, the sample is shorter.[47] We therefore use the 13F data as another opportunity to try to estimate the elasticity using a different level of aggregation.

We consider an extension of the model in (40), where we allow for heterogeneity in demand elasticities, $\zeta_{j,t-1}$:

$$\Delta q_{jt} = -\zeta_{j,t-1} \Delta p_t + \lambda_{j,t-1}' \eta_t + u_{jt}.$$

[46] In the US, all institutional investment managers managing over $100 million or more in "13F securities" (which include stocks) must report their holdings on Form 13F every quarter.

[47] Alternatively, we can use data from Thomson Reuters that has a history going back to 1980. However, Thomson Reuters has several measurement issues that can lead to some large flows, which is potentially problematic for the GIV estimator that exploits large, idiosyncratic shocks.

We assume a parametric specification for elasticities and a semi-parametric specification for factor loadings:

$$\zeta_{j,t-1} = \dot{\zeta}' x_{j,t-1}, \qquad \lambda_{j,t-1} = \dot{\lambda}' x_{j,t-1} + \ddot{\lambda}_j,$$

where $x_{j,t-1}$ is a vector of investor characteristics of which the first element is equal to 1, and $\dot{\zeta}$, $\dot{\lambda}$, and $\ddot{\lambda}_j$ are to be estimated. As investor characteristics, we use log turnover, active share, and log AUM, as well as portfolio characteristics, namely the value-weighted market capitalization, log book-to-market ratio, and market beta, to estimate the factors.[48] We also provide further details on measuring flows, sample selection criteria, and how we construct the characteristics in Appendix C.3. The macro elasticities can also be heterogeneous across investors and depend on the same, or a subset of the, characteristics. In addition, we allow for non-parametric factors via $\ddot{\lambda}_j$, as before.

Using this structure, we implement a version of the GIV procedure, as in Gabaix and Koijen (2020), that we detail in Section B.2. The results are presented in Panel A of Table 2. In the first column, we only use parametric factors, that is, $\ddot{\lambda}_j = 0$. We estimate a multiplier of $M = 6.0$ with a standard error of 3.3. In columns 2 to 6, we add one to five principal components, respectively. This is in addition to six parametric factors associated with the characteristics and a common factor on which all investors load equally. The multiplier varies between $M = 3.8$ and $M = 5.3$. The standard errors tighten as we add the principal components and we explain more of the variation, as can be seen from the R-squared value that increases from 23% to 77%. The increase in the R-squared value flattens after two principal components.

To conclude, we also implement an over-identifying restrictions test as suggested by Gabaix and Koijen (2020). After estimating $\Delta\check{q}_{jt}$, we construct two GIV instruments. Each period, we rank all funds by their lagged size (large to small) and we put the funds with an odd rank in one group and those with an even rank in the other group. We then construct a separate instrument for the odd group, Z_t^{Odd}, and for the even group, Z_t^{Even}. By design, $Z_t = Z_t^{Odd} + Z_t^{Even}$. We also construct their difference, $Z_t^{Odd} - Z_t^{Even}$.

We then run the final regression

$$\Delta p_t = \alpha + M Z_t + M^{Odd-Even}\left(Z_t^{Odd} - Z_t^{Even}\right) + \lambda_P' \eta_t^e + \epsilon_t,$$

where $\eta_t^e = \left(\Delta y_t, \eta_t^{PC,e}\right)$ and $\eta_t^{PC,e}$ are the factors that we estimate using principal components. We then test whether $H_0 : M^{Odd-Even} = 0$. The results are presented in Panel B of Table 2. We find that the estimates of M are close to those in Panel A and the coefficient on the difference is statistically insignificant. Of course, given the fairly short sample, we do emphasize that the standard errors are quite large throughout.[49]

As discussed before, the main concern with GIV estimates is that we accidentally omit a factor whose loadings are correlated with size. To alleviate this concern, we control for log size in the parametric loadings and extract up to five principal components. Another test is to explore the correlation between Z_t^{Odd} and Z_t^{Even}. As we split funds into odd and even after ranking them by size, a size-related factor may be present in both Z_t^{Odd} and Z_t^{Even} and would introduce a correlation between the series. In Panel C of Table 2, we present the results of regressing Z_t^{Odd} on Z_t^{Even}. In all cases, we find that the correlations are insignificant and the R-squared values are very small.

[48] In terms of the characteristics, it is important to control for log size, as also discussed in Gabaix and Koijen (2020). Without properly controlling for factors whose loadings depend on size, the multiplier estimates can be biased.

[49] Also, the concentration using 13F institutions is less pronounced than using the FoF data, which leads to larger standard errors in the GIV procedure.

Table 2: Multiplier estimates using 13F data. *Panel A*. We report the multiplier estimates using 13F data. In the first column, we only use parametric factors. The characteristics used are log AUM, active share, log turnover as well as the value-weighted size of the portfolio, log market-to-book ratio, and market beta. In the next five columns we add one to five principal components, respectively. All regressions include a constant and control for GDP growth and the principal components, which we omit for brevity. *Panel B*. We test for over-identifying restrictions by constructing two GIV instruments. We sort funds each period by their lagged size and split them into an odd and even group based on their ranks. The sum of the instruments is the instrument used in Panel A, and we regress returns on Z, the difference in instruments, $Z^{Odd} - Z^{Even}$, and the standard controls (GDP growth and up to five principal components). Under the null of the model, the coefficient on Z_t should be the same as in Panel A and the coefficient on $Z^{Odd-Even}$ should be zero. *Panel C*. We regress the instrument based on the odd observations on the instrument based on the even observations, while adding the same controls as before. Under the null of the model, the regression coefficient should be zero. We report the standard errors, which are corrected for autocorrelation, in parentheses. The sample is from 1999.Q2 to 2017.Q2.

Panel A: Benchmark results

Number of principal components	0	1	2	3	4	5
Z	5.97	5.34	3.93	3.83	4.19	4.05
	(3.28)	(1.79)	(2.36)	(1.72)	(1.68)	(1.62)
Observations	72	72	72	72	72	72
R^2	0.233	0.545	0.751	0.752	0.759	0.766

Panel B: Over-identifying restrictions test

Number of principal components	0	1	2	3	4	5
Z	5.72	5.23	3.93	3.81	4.17	4.04
	(3.19)	(1.69)	(2.35)	(1.68)	(1.63)	(1.57)
$Z^{Odd} - Z^{Even}$	1.65	0.73	-0.03	0.11	0.12	0.06
	(3.64)	(1.43)	(1.06)	(0.99)	(0.72)	(0.73)
Observations	72	72	72	72	72	72
R^2	0.237	0.545	0.751	0.752	0.759	0.766

Panel C: Regression of Z^{Odd} on Z^{Even}

Number of principal components	0	1	2	3	4	5
Z^{Even}	-0.13	-0.13	-0.13	-0.13	-0.14	-0.14
	(0.14)	(0.14)	(0.14)	(0.12)	(0.15)	(0.15)
Observations	72	72	72	72	72	72
R^2	0.018	0.022	0.036	0.036	0.048	0.051

To provide some further insight in the variation of the instrument, $Z_t = \sum_j S_{j,t-1} u_{jt}$, we can inspect the largest shocks, $S_{j,t-1} |u_{jt}|$. As is to be expected, the largest shocks are driven by investors such as State Street Global Advisors, BlackRock, et cetera. Unfortunately, however, it is typically not possible to provide a simple narrative as to why a shock occurs in a given period using 13F data as the filings reflect the aggregate positions across a variety of funds and accounts.

In summary, we find that the multiplier estimates are quite consistent with the estimates we found using the FoF data.

4.4 Which investor sectors make the market fluctuate the most?

Given the estimates of the elasticities, we can decompose the relative contributions of different sectors to stock market fluctuations as we illustrate using Flow of Funds data. We first compute $f^{\nu}_{jt} = \Delta q_{jt} + \zeta \Delta p_t$, which is the combined impact of capital flows and demand shocks. For the corporate sector, given the low elasticity that we estimate at a quarterly frequency, we set $f^{\nu}_{Ct} = \Delta q_{Ct}$. The contribution of sector j to the overall variance is

$$s^p_j = \frac{Cov\left(\Delta p_t, \frac{1}{\zeta} S_{j,t-1} f^{\nu}_{jt}\right)}{Var\left(\Delta p_t\right)}, \tag{42}$$

which sums to 100% across investors as we include the corporate sector, where $\frac{1}{\zeta} S_{0,t-1} f^{\nu}_{0t} = \frac{1}{\zeta} f^{\nu}_{Ct}$. We report the variance shares in Figure 4 based on the third column of Table 1. We find that the household sector accounts for 54% of the variation in stock returns, followed by mutual funds and ETFs (30%), the foreign sector (14%), and private pension funds (7%). The somewhat negative contribution of state and local pension funds may reflect once again their somewhat higher demand elasticity. The share of variance is different from the equity share presented in Figure 1, so the variance shares are not solely driven by size.

Next, we study the properties of sector-specific shocks in more detail. We run the first three steps of the GIV algorithm outlined in Section B.1 and estimate u_{it} as the residuals of regressing $\Delta \check{q}_{it}$ on η^e_t. In Table D.11 in the Online Appendix, we report the volatility of sector-specific demand shocks by sector.

An important feature of stock return data is that the volatility changes over time. We find that sector-specific demand shocks are heteroscedastic as well. To see this, we compute the two-year rolling variance of sector-specific idiosyncratic shocks, u_{jt}, and refer to it as $\sigma^2_{u_j,t}$. We then compute $\sigma_{u_{St}} = \left(\sum_j S^2_{j,t-1} \sigma^2_{u_j,t}\right)^{\frac{1}{2}}$ and we also compute the rolling standard deviation for stock returns.

Figure 5 plots both series (both standardized), which are positively correlated. Their time series correlation is 41%. During bad times, aggregate shocks and sector-specific shocks both increase in terms of their volatility. In future work, it would be interesting to explore the micro-foundation of heteroscedasticity in demand shocks in greater detail.

4.5 Some descriptive statistics on flows and their possible determinants

In the final stretch of this empirical section, we provide descriptive statistics on flows. We hasten to say that here, unlike in the rest of this Section 4, we loosen our identification standards and show only correlations, without an attempt at causal inference. We still find those correlations suggestive, and hope that they will inspire researchers to dig deeper.

Figure 4: Who moves the market? The figure reports the contribution to the variance of the US equity market of different sectors. The variance shares are computed using equation (42). This may be compared to the "size shares" in Figure 1.

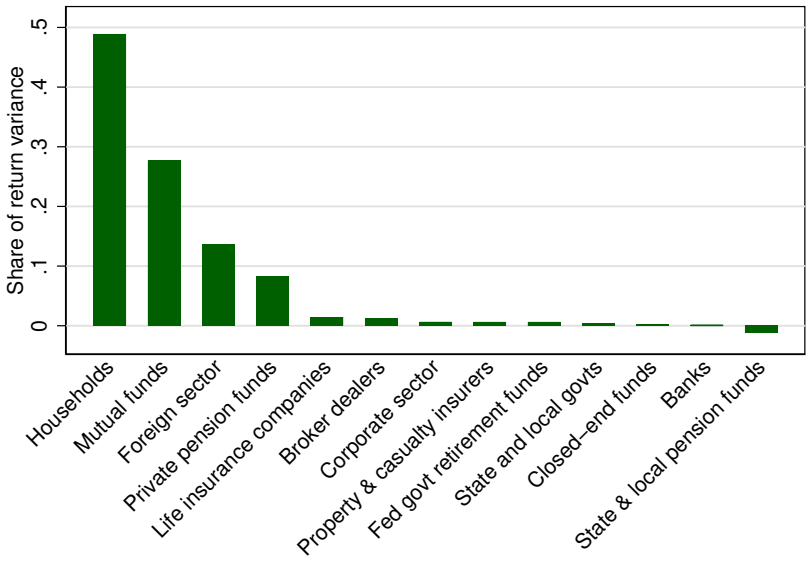
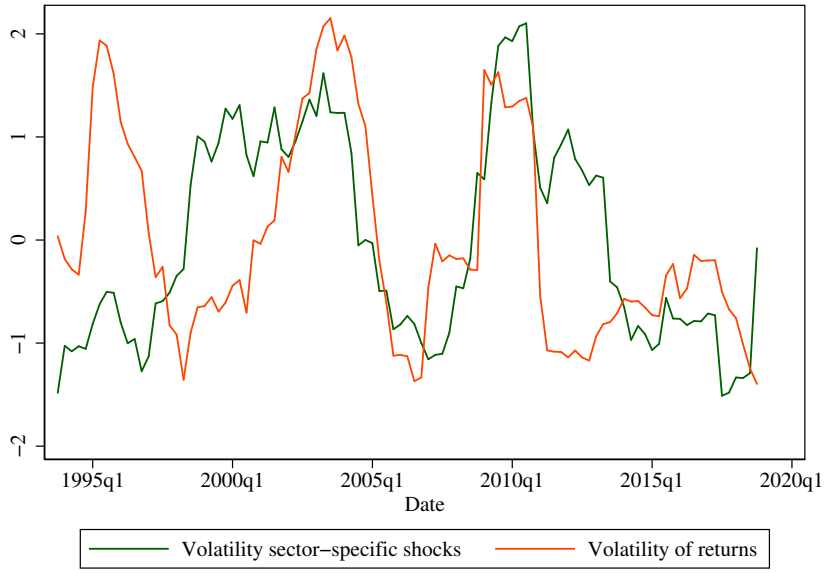


Figure 5: Volatility of flows and volatility of returns. We compute the volatility of idiosyncratic demand shocks by the various sectors (u_{St}) using a two-year rolling average and we do the same for the return on the aggregate stock market. Both series are standardized. The sample is from 1994.Q1 to 2018.Q4.



The relative importance of capital flows versus demand shocks So far, we have studied f^{ν}_{jt}, which is the sum of capital flows, f_{jt}, and demand shocks, ν_{jt}. In this section, we want to know their respective importance, that is, we want to know

$$\Phi = \frac{Cov\left(f_{jt}, f^{\nu}_{jt}\right)}{Var\left(f^{\nu}_{jt}\right)}, \tag{43}$$

so that Φ is the share of the variance that is due to flows, and $1-\Phi$ the share due to demand shocks (the decomposition is motivated by the identify $Var\left(f^{\nu}_{jt}\right) = Cov\left(f_{jt}, f^{\nu}_{jt}\right) + Cov\left(\nu_{jt}, f^{\nu}_{jt}\right)$).

We explore Φ using the fund-level data for mutual funds, ETFs, and state and local pension funds. To estimate ν_{jt}, we run the regression (where we expect $\beta \simeq 1$)

$$\Delta q_{jt} = \alpha_j + \gamma \Delta p_t + \beta f_{jt} + \nu_{jt}, \tag{44}$$

then collect the residuals, ν_{jt} and compute $f^{\nu}_{jt} = f_{jt} + \nu_{jt}$. Then, we estimate Φ by the regression $f_{jt} = \alpha_j + \Phi f^{\nu}_{jt} + u_{jt}$.

We find $\Phi = 77\%$ (s.e. 0.01) for mutual funds and $\Phi = 89\%$ (s.e. 0.02) for ETFs.[50] These estimates are likely to be lower bounds as some measurement error is introduced by using the aggregate market return as opposed to the return on the equity portion of the portfolio.[51] Intuitively, the vast majority of ETFs are close to all-equity or fixed share strategies and $f_{jt} = f^{\nu}_{jt}$, by definition. For state and local pension funds, we estimate $\Phi = 37\%$ (s.e. 0.05).[52] This share is high considering that the inflows into state and local pension plans are fairly stable (that is, the inflows are typically proportional to government salaries and the payouts to plan participants are highly predictable). Based on these simple calculations, we estimate that capital flows may account for more than half of the variation in f^{ν}_{jt}.

A strong limitation of this subsection is that we can do this decomposition for only some segments of the market (mutual funds, ETFs and state and local pension funds). Still, this leads to the provisional estimate than the majority of changes in the demand for equities ($f^{\nu}_{jt} = f_{jt} + \nu_{jt}$) is from capital flows from investors (f_{jt}), rather than from demand shocks by fund managers (ν_{jt}).

A measure of aggregate capital flows into the stock market To connect capital flows to prices and macro variables, we need to measure them ideally for each sector. So we return to the FoF data and construct an estimate of capital flows into the equity market in a way that is consistent with our theory. As this measure is new to the literature, we show its connection to prices, macro-economic variables, and beliefs. The results in this section provide an initial analysis of the potential determinants of flows into the aggregate stock market.

---

[50] We use the aggregate market return for each fund in (44). Also, for Δq_{jt}, we use $\Delta q_{jt} = \frac{A_{jt}\theta_{jt} - A_{j,t-1}\theta_{j,t-1}(1+r_t)}{A_{j,t-1}\theta_{j,t-1}}$, where r_t is once again the aggregate market return. Appendix D.4 provide details.

[51] It may be possible to refine these calculations using the equity holdings data for mutual funds to get the fund specific equity return.

[52] For state and local pension funds, we define the dollar flow as $F^i_{jt} = A_{jt}\theta^i_{jt} - A_{j,t-1}\theta^i_{j,t-1}\left(1 + r^i_{jt}\right)$, where r^i_{jt} is investor- and asset-class specific and $i = E, B$ for equity and fixed income. We then compute $f_{jt} = \frac{F^E_{jt} + F^B_{jt}}{A_{j,t-1}\left(\theta^E_{j,t-1} + \theta^B_{j,t-1}\right)}$ and $\Delta q_{jt} = \frac{F^E_{jt}}{A_{j,t-1}\theta^E_{j,t-1}}$. In the case of pension funds, as we have the fund-specific equity returns, we estimate (44) using these fund-specific equity returns as r_t. For all three data sets, that is, for pension funds, mutual funds, and ETFs, we only include funds with at least 15 observations and we winsorize Δq_{jt} and f_{jt} at the 5% and 95% level per period.

We rely on the FoF for these calculations, and we refer to Appendix C for details on the data construction of the fixed income positions and flows. However, the FoF aggregates data across many institutions, and the reported flows can be mismeasured. To see this, consider the case in which some households only invest in bonds and other households only invest in equities. If we view this as a combined household, a 1% combined inflow into financial markets does not necessarily lead to a 1% increase in equity holdings as the flow may be a flow to bond funds only.[53] With disaggregated data, such problems can be solved, but such data are unfortunately unavailable.

We propose a simple diagnostic to assess whether flows are measured accurately. In particular, in our model, the elasticity of demand to flows equals one, see (7). We therefore estimate

$$\Delta q_{jt} = \alpha + \beta_j f_{jt} + \gamma_j \Delta p_t + \delta_j' x_{jt} + \epsilon_{jt}, \tag{45}$$

where x_{jt} includes $f_{j,t-1}$ and Δy_t. We only include the flows for which we cannot reject the null hypothesis $H_0 : \beta_j = 1$. We report the estimates of β_j in Table D.12 in Appendix D. When we cannot reject $H_0 : \beta_j = 1$, we use the total flow and for the other sectors we replace it with the equity flow. We refer to these "screened flows" as f^*_{jt}. [54]

The correlation between capital flows and equity returns We relate our measure of capital flows into the stock market to returns. In the left panel of Figure 6, we show that our measure of flows is strongly correlated with returns using a binned scatter plot in the left panel. We again find that the slope is high, but we emphasize that because of endogeneity that slope is not a good measure of the impact of flows of the price, which is why earlier we developed an IV strategy to measure that impact.[55]

We can also illustrate the strong co-movement between flows and prices at lower frequencies. In particular, we construct a cumulative (log) return index and compute cumulative flows. We then extract the cyclical component using the methodology developed in Hamilton (2018). We standardize both measures over the full sample to be able to plot them in the same figure. These are shown in the right panel of Figure 6. Consistent with the high-frequency co-movement that we uncover in the left panel of Figure 6, we find that prices and flows co-move at a business cycle frequency. We re-emphasize once again that these are merely correlations and it may be the case, for example, that they reflect positive feedback trading by investors (Cutler et al. (1990), Shleifer (2000)).

Relating flows to shocks to GDP and to return expectations To conclude this initial exploration of capital flows into equity markets, we relate flows to shocks to economic activity and survey expectations of returns. We use GDP growth as our measure of economic activity, as before. For return expectations, we use the survey from Gallup. The data are described in more detail in Appendix C. Gallup has several missing observations and only starts in 1996.Q4. We only

[53] Conceptually, within say the household sector, we want to total flow weighing by the equity share of each component (see Section 3.3), but the FoF provides only the total flow weighing by wealth weights.

[54] Alternatively, we can use a cutoff of $|\beta_j - 1| < 0.3$, for instance. The cutoff of 0.3 is obviously somewhat arbitrary and changes in this cutoff do not matter much (e.g., 0.2 or 0.4) until the household sector is included (the household sector has a slope coefficient of $\beta_j = 0.45$, see Table D.12). This sector is large and flows do not explain equity purchases well.

[55] If one has data on capital flows for a substantial number of sectors, then it would be possible to construct a GIV estimate based on capital flows alone. This would make it possible to estimate the causal impact of prices on capital flows, and of capital flows on prices.

Figure 6: Capital flows into the stock market and price changes. We plot the aggregate flow into the stock market, using the screened flows, f^*_{jt}, $f^*_{St} = \sum_{j=0}^{N} S_j f^*_{jt}$, versus the return on the aggregate stock market in the left panel used a binned scatter plot. In the right panel, we construct a cumulative (log) return index and compute cumulative flows. We extract the cyclical component using the methodology developed in Hamilton (2018). We standardize both measures over the full sample to be able to plot them in the same figure. The sample for both figures is from 1993.Q1 to 2018.Q4.

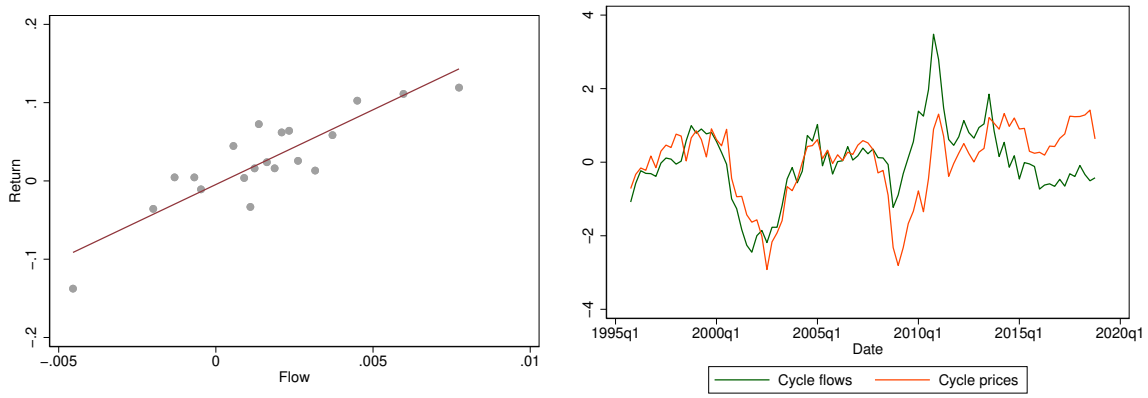


Table 3: Descriptive statistics on capital flows, survey expectations of beliefs, economic activity, and stock returns. The table reports the time-series regressions of innovations to flows in the first three columns on innovations to survey expectations of returns (column 1), GDP growth innovations (column 2), and both variables combined (column 3). We estimate the innovations in all cases by estimating an AR(1) model, and normalize them to have unit standard deviation. Then we regress returns on flow innovations (column 4), innovations to survey expectations of returns (column 5), GDP growth innovations (column 6), and all three variables combined (column 7). The sample is from 1997.Q1 to 2018.Q4, with some gaps, due to missing data for the Gallup survey.

	Flow	Flow	Flow	Return	Return	Return	Return
Gallup	0.48		0.46		0.61		0.33
	(0.10)		(0.11)		(0.09)		(0.09)
GDP growth		0.21	0.06			0.41	0.21
		(0.11)	(0.11)			(0.10)	(0.08)
Flow				0.65			0.45
				(0.09)			(0.09)
Constant	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.10)	(0.11)	(0.10)	(0.09)	(0.09)	(0.10)	(0.07)
Observations	79	79	79	79	79	79	79
R^2	0.233	0.046	0.237	0.426	0.376	0.171	0.582

use data for all series when they are non-missing, which gives us 79 quarterly observations. To obtain innovations, we estimate an AR(1) model for each of the series. We standardize each of the innovation series, by removing the time-series mean and dividing it by their standard deviations, to simplify the interpretations of the regressions.

The results are presented in Table 3. In the first three columns, we relate capital flows to survey expectations and economic growth. We find that flows and survey expectations are strongly correlated, confirming Greenwood and Shleifer (2014) using a more comprehensive measure of capital flows. A one standard deviation increase in survey expectations of returns is associated with a 0.48 standard deviation increase in capital flows.

This finding may point to a resolution of a recent challenge posed to the beliefs literature by Giglio et al. (2019). In particular, they find that although survey expectations of returns are volatile, the pass-through to actions (that is, portfolio rebalancing) is low. One possibility is that the strong correlation between innovations to beliefs and prices (which equals 61% in our sample) arises even though the pass-through is low, but small flows into inelastic markets lead to large price effects.

Flows and economic activity, as analyzed in the second column, are also positively correlated, but the relation is substantially weaker. In the third column, we combine survey expectations and economic activity, and find that the latter is insignificant. In the remaining columns, we study the association between returns and flows, beliefs, and economic activity. A one standard deviation increase in capital flows is associated with a 0.65 standard deviation increase in returns, which is similar to a 0.61 standard increase in case of survey expectations. The link to GDP growth is significant, but weaker with a slope coefficient of 0.41. In the final column, we combine all flows, beliefs, and GDP growth and find that even in this multiple regression, all variables are significant. The R-squared of this final regression is high and amounts to $R^2 = 58\%$.

Obviously, this analysis is just an initial exploration into the determinants of flows, and more disaggregated data may be used to explore the determinants of capital flows for various institutions and across households. If the inelastic markets hypothesis holds, this is an important area for future research.

5 General Equilibrium with Inelastic Markets

So far, we took both the risk-free rate r_f and the average equity premium $\bar{\pi}$ as exogenous. We now endogenize them. For instance, we shall see how flows from bonds to stocks, which alter the price of stocks, can at the same time keep the risk-free rate constant (in our model, this is because the optimizing household also trades off saving in bonds versus consumption, and this way ensures that the consumption-based Euler equation for bonds holds). We view this as a prototype for how to build general equilibrium models with inelastic markets, merging behavioral disturbances, the flows they create, their impact on prices, and potentially their impact on production.

5.1 Setup

We consider an infinite horizon economy. For simplicity, we discuss in detail an endowment economy – then, it will be easy to generalize to a production economy. The endowment Y_t follows a proportional growth process, with an i.i.d. lognormal growth rate G_t:

$$Y_t = G_t Y_{t-1}, \qquad G_t = e^{g+\varepsilon_t^y - \frac{1}{2}\sigma_y^2}$$

with $\varepsilon_{t+1}^{y} \sim \mathcal{N}\left(0, \sigma_y^2\right)$. Utility is $\sum_t \beta^t u\left(C_t\right)$ with $u\left(C\right) = \frac{C^{1-\gamma}}{1-\gamma}$.

Because empirically dividends and GDP are not very correlated, we model that GDP Y_t is divided as $Y_t = \mathcal{D}_t + \Omega_t$ into an aggregate dividend $\mathcal{D}_t$ and a residual Ω_t, where the dividend stream has i.i.d. lognormal growth, $\mathcal{D}_t = G_t^D \mathcal{D}_{t-1}$, $G_t^D = e^{g + \varepsilon_t^D - \frac{1}{2}\sigma_D^2}$. The "residual" Ω_t can be thought of as a combination of wages, entrepreneurial income, and so forth (and indeed is the majority of GDP).[56] The representative firm raises capital entirely through equity, and passes the endowment stream as a dividend, $D_t = \frac{\mathcal{D}_t}{Q^{\mathcal{E}}}$, where $Q^{\mathcal{E}}$ is the number of shares of equities supplied by the corporate sector, which is an unimportant constant in this baseline model without share buybacks and issuances. Bonds are in zero net supply.[57] We write the price of equities as

$$P_t = \frac{D_t}{\delta} e^{p_t}, \tag{46}$$

where δ is the average dividend-price ratio and p_t (with $\mathbb{E}\left[e^{p_t}\right] = 1$) is the deviation of the price from the baseline $p_t = 0$. Those quantities are all endogenous.

There are two funds: a pure bond fund, which just holds bonds, and the mixed fund, which holds bonds and equities. The mixed fund has a mandate, to hold a fraction in equities equal to:

$$\theta_t = \theta \exp\left(-\kappa^D p_t + \kappa \mathbb{E}_t\left[\Delta p_{t+1}\right]\right), \tag{47}$$

which is the same as before in (3), to the leading order (in terms of deviations from the steady state), with $\kappa^D = \kappa\delta$. The formulation here is slightly more general.[58]

Consumption and investment by households We describe the behavior of the representative household. Section E.9 supplies more formalism and further details. The dynamic budget constraint of household h entails:[59]

$$Q_t^{B,h} + D_t^h + \Omega_t^h = C_t^h + \Delta F_t^h + \frac{Q_{t+1}^{B,h}}{R_{f,t}}. \tag{48}$$

Indeed, the left-hand side is the bond asset position of the households at the beginning of period t: $Q_t^{B,h}$ gives the bond holdings at the beginning of period t, while D_t^h and Ω_t^h are the dividend and residual income received by the household in its pure bond fund (which includes the "dividends" paid by the mixed fund). This bond position is spent on consumption C_t^h, flows ΔF_t^h into the mixed fund fund, and investment in bonds, with a face value $Q_{t+1}^{B,h}$.

We need a behavioral element, otherwise the investor would fully undo the funds' mandate. We choose to decompose the household as a rational consumer, who only decides on consumption (so dissaving from the pure bond fund), and a behavioral investor, who trades between the pure bond fund and the mixed fund.

[56] Formally, it could become negative, though this is a very low probability event in our calibration. Then, the interpretation is that of a residual liability. In addition, it would be easy to keep $\mathcal{D}_t/Y_t$ stationary, at the cost of having it as one more state variable, reverting to its mean.

[57] We could have the government issue bonds, backed by taxation: the government would issue $Q_t^B = \xi Y_t$, for a constant ξ (the tax at time t ensures budget balance, so $\mathcal{T}_t = -Q_t^B + R_{f,t-1} Q_{t-1}^B$). Likewise, corporations could also issue bonds. Then, the same economics would hold, only with extra notations.

[58] But here we allow the mandate to potentially differentiate between "return predictability coming from the price-dividend ratio" (captured by $-\kappa^D p_t$) and "return predictability because the price is predictable". In a number of settings the first one (the "carry") is stronger than the last one (Koijen et al. (2018)), so having two κ's is sensible.

[59] There is also the usual transversality condition, $\lim_{t\to\infty} \beta^t \left(C_t^h\right)^{-\gamma} Q_t^{B,h} = 0$.

The rational consumer part of the household maximizes lifetime utility, subject to the dynamic budget constraint for bonds (48). She takes the actions of the investor as given.[60] As she is rational, she satisfies the Euler equation for bonds:

$$\mathbb{E}_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}R_{ft}\right]=1, \tag{49}$$

with $C_t = Y_t$ in equilibrium. This pins down the interest rate R_{ft}, which is constant in our i.i.d. growth economy.

The behavioral investor part of the household is influenced by b_t, a behavioral disturbance. It is a simple stand-in for noise in institutions, beliefs, tastes, fears, and so on. We assume that the investor trades (between stocks and bonds) with a form of "narrow framing" objective function. He seeks to maximize $\mathbb{E}_t\left[V^p\left(W_{t+1}\right)\right]$ with $V^p\left(W\right)=\frac{W^{1-\gamma}}{1-\gamma}$ a proxy value function. Specifically, when $b_t = 0$, he chooses his allocation θ^M in the mixed fund as:

$$\theta^{M*}=\underset{\theta^M}{\operatorname{argmax}}\,\mathbb{E}\left[V^p\left(\left(1-\theta^M\right)R_{ft}+\theta^M R_{M,t+1}\right)|b_t=0\right], \tag{50}$$

where $R_{M,t+1}$ is the stochastic rate of return of the mixed fund. This choice of a "narrow framing" benchmark is opposed to the full rational value function, which would have all the Merton hedging demand terms. The latter would lead to the consumption CAPM holding on average: in particular, the equity premium would be too small (at $\bar{\pi}=\gamma\sigma_y^2$). Instead, the above formulation will lead to $\bar{\pi}=\gamma\sigma_r^2$, where the σ_r^2 is the volatility of the stock market.[61]

If there are no behavioral disturbances, an investor wishing to maintain a constant allocation θ^{M*} in the mixed fund should invest into it via the following flow, with $\mathcal{D}^t \coloneqq (\mathcal{D}_t, \mathcal{D}_{t-1})$:[62]

$$\Delta F^*\left(\mathcal{D}^t\right)=\frac{1-\theta}{\theta}\Delta\left(\frac{\mathcal{D}_t}{\delta}\right). \tag{51}$$

We assume that his policy, however, is affected by the behavioral disturbance b_t, so that the actual flow is

$$\Delta F\left(\mathcal{D}^t,b_t\right)=\Delta F^*\left(\mathcal{D}^t\right)+\Delta\left(b_t\frac{\mathcal{D}_t}{\delta}\right). \tag{52}$$

i.e. is higher than the baseline amount $\Delta F^*\left(\mathcal{D}^t\right)$ (where there are no behavioral disturbances) by a fraction Δb_t of the "fundamental value" $\frac{\mathcal{D}_t}{\delta}$ of the equity market — i.e., of its value evaluated at the typical dividend price ratio δ. This means that the flow in the mixed fund is Here we will specify that b_t is an AR(1) with speed of mean-reversion ϕ_b. The state vector is

$$Z_t=\left(Y_t,\mathcal{D}^t,b_t\right), \tag{53}$$

where $(Y_t, \mathcal{D}^t)$ are the fundamentals, and b_t the behavioral deviation.

[60] One could imagine a variant, where the consumer manipulates the investor's actions. This would lead her to distort her Euler equation for consumption.

[61] This choice of "narrow framing" leads to a high equity premium. It could be replaced by another device, e.g. disasters. We choose here narrow framing as this behavioral ingredient is well in the behavioral spirit of this section.

[62] Indeed, consider the economy in the undisturbed state, $b_t = 0$. In that state, the aggregate value of equities is $\mathcal{P}^*\left(\mathcal{D}_t\right)=\frac{\mathcal{D}_t}{\delta}$ (remember that $\mathcal{D}_t$ is the aggregate dividend) and the value of the mixed fund is $W^*\left(\mathcal{D}_t\right)=\frac{1}{\theta}\mathcal{P}^*\left(\mathcal{D}_t\right)$, and it should hold $(1-\theta)W^*\left(\mathcal{D}_t\right)$ quantities of bonds. Hence, the flow to the mixed fund $\Delta F^*\left(D^t\right)=(1-\theta)\Delta W^*\left(\mathcal{D}_t\right)=\frac{1-\theta}{\theta\delta}\Delta\mathcal{D}_t$.

Definition 1. An equilibrium comprises the following functions: the stock-price $P(Z)$, the interest rate $R_f(Z)$, and the consumption and asset allocation $C(Z)$, $B(Z)$, such that the mixed fund's allocation $\theta(P, Z)$ follows its mandate, and:

1. The consumer follows the consumption policy $C(Z)$, which maximizes utility subject to the above constraints.
2. The investor follows the behavioral policy (52), where the average allocation in the mixed fund is given by (50), so that it is quasi-rational with narrow framing on average, but with disturbance b_t.
3. The mixed fund's demand for stock $Q^D(Z)$ follows its mandate (47).
4. The consumption market clears, $C(Z) = Y(Z)$.
5. The equity market clears: the mixed fund holds all the equity, so that $Q^D(Z) = Q^{\mathcal{E}}$.

5.2 Model solution

Proposition 7 describes the solution of that economy. In particular, it shows that the link between the disturbance b_t and the cumulative flow $\tilde{f}_t$ is as follows. We start from an equilibrium situation, where $b_0 = 0$. The cumulative flow since time 0 can be written, to the leading order in b_t and d_t:

$$f_t = (1-\theta)\, d_t + \tilde{f}_t, \qquad \tilde{f}_t = \theta b_t. \tag{54}$$

where $d_t = \sum_{s=1}^{t} \frac{\Delta D_s}{D_{s-1}}$ is the cumulative growth rate in the dividend. This holds for any process b_t. Now, we specialize to the case where b_t follows an AR(1). Then, so does $\tilde{f}_t$, as in

$$\tilde{f}_t = (1-\phi_f)\, \tilde{f}_{t-1} + \varepsilon_t^f, \tag{55}$$

so that we are in the "convenient benchmark" case of (21)-(22). But in this general equilibrium section, we now endogenize the interest rate and the average equity premium $\bar{\pi}$.

This AR(1) assumption is just a placeholder for richer behavioral assumptions, for example driven by time-varying beliefs (as in Caballero and Simsek (2019), Bordalo et al. (2020)), positive or negative feedback trading rules, and so on. We seek here only a coherent, simplified model, which can be fully solved and which lends itself to a number of variants. Importantly, it relies on observables flows, which reduces the need for "dark matter" in asset pricing.

Proposition 7. *The solution of the economy obtains in closed form as follows, taking the limit of small time intervals and only the first order terms in $\tilde{f}_t$. The market elasticity ζ and its "macro market effective discount rate" ρ (see Proposition 5) are:*

$$\zeta = 1 - \theta + \kappa^D, \qquad \rho = \frac{\zeta}{\kappa}. \tag{56}$$

The price of equities is

$$P_t = \frac{D_t}{\delta} e^{p_t}, \tag{57}$$

where D_t is the dividend, $\delta = r_f + \bar{\pi} - g$ is the average dividend-price ratio, and p_t is the deviation of the price from its rational average, which increases with flows:

$$p_t = b_f^p \tilde{f}_t, \qquad b_f^p = \frac{1}{\zeta + \kappa \phi_f}. \tag{58}$$

Hence the variance of stock market returns is

$$\sigma_r^2 = var\left(\varepsilon_t^D + b_f^p \varepsilon_t^f\right), \tag{59}$$

and depends on both fundamental risk (ε_t^D) and flow-risk (ε_t^f). Both contribute to the average equity premium, which is

$$\bar{\pi} = \gamma \sigma_r^2. \tag{60}$$

The equity premium at time t is lower than its average when flows have been high, as:

$$\pi_t = \bar{\pi} + b_f^\pi \tilde{f}_t, \qquad b_f^\pi = -\left(\delta + \phi_f\right) b_f^p. \tag{61}$$

Finally, the interest rate is constant, and given by the consumption Euler equation (49)*:*

$$r_f = -\ln \beta + \gamma g - \gamma\left(\gamma + 1\right) \frac{\sigma_y^2}{2}. \tag{62}$$

Hence, we have a fairly traditional economy, except that, crucially, prices and risk premia are now driven by flows and flow risk, in addition to fundamentals. Hence, the equity premium is time-varying (because of flows), and on average higher than in the consumption CAPM (because it reacts to flow risk, not just fundamental risk), as given in (60).

5.3 Pricing kernel consistent with flow-based pricing

We show how to express the economics of flows in inelastic markets in the language of pricing kernels or stochastic discount factors (SDFs). To do so, we use a simple general method to complete a "default" pricing kernel so that it reflects the impact of flows on asset prices. The idea is simply that there is a fringe of infinitesimal traders that can absorb any infinitesimal amount of new assets. That gives rise to a "completed" pricing kernel (see Section E.10 for details). In our general equilibrium model, the completed SDF is:

$$\mathcal{M}_{t+1} = \exp\left(-r_f - \pi_t \frac{\varepsilon_{t+1}^D}{\sigma_D^2} + \xi_t\right), \qquad \pi_t = \bar{\pi} + b_f^\pi \tilde{f}_t, \tag{63}$$

where $\sigma_D^2 = var\left(\varepsilon_{t+1}^D\right)$ and ξ_t is a deterministic term ensuring that $\mathbb{E}_t\left[\mathcal{M}_{t+1}\right] e^{r_f} = 1$, so that $\xi_t = -\frac{\pi_t^2}{2\sigma_D^2}$ if ε_{t+1}^D is Gaussian.[63]

[63] Likewise, in the two-period model of Section 3.1, the excess equity premium is $\hat{\pi} = \delta\left(d - p\right)$ with p given in (10), so that, with $f = \left(1 - \theta\right) d + \tilde{f}$, the total equity premium is: $\pi = \bar{\pi} - \delta \frac{\tilde{f}}{\zeta}$. So, the completed pricing kernel is:

$$\mathcal{M} = \exp\left(-r_f - \pi \frac{\varepsilon^D}{\sigma_D^2} + \xi\right), \qquad \pi = \bar{\pi} - \delta \frac{\tilde{f}}{\zeta}, \tag{64}$$

with $\xi = -\frac{\pi^2}{2\sigma_d^2}$ if ε^D is Gaussian. This SDF prices correctly stocks and bonds. Section E.8 develops the general equilibrium model in a two-period economy.

This gives the "flow-based" completed pricing kernel, which is an alternative to the consumption-based kernel of Lucas (1978). The core economics is in how flows affect prices, and the pricing kernel (63) just reflects that. As there is a flow $\tilde{f}_t$ that modifies the price P_t according to Proposition 7, and the pricing kernel $\mathcal{M}_{t+1}$, in such a way that $P_t = \mathbb{E}_t\left[\mathcal{M}_{t+1}\left(D_{t+1} + P_{t+1}\right)\right]$ holds. The pricing kernel is in a sense a symptom rather than a cause in that market.

5.4 Production: Basic equations of macro-finance with flows

We now recap how the basic equation of macro-finance work with flows. In the general equilibrium model above consumption prices the risk-free rate, but it does *not* price equities. Generically, we have that

$$\mathbb{E}_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R_{ft}\right] = 1, \qquad \mathbb{E}_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R_{M,t+1}\right] \neq 1.$$

Instead, the SDF (63) reacts to flows, and prices the risk-free rate and equities:

$$\mathbb{E}_t\left[\mathcal{M}_{t+1} R_{ft}\right] = 1, \qquad \mathbb{E}_t\left[\mathcal{M}_{t+1} R_{M,t+1}\right] = 1.$$

If we had a production-based model with capital K_t, then investment I_t and labor demand L_t (with κ the cost of investment, w the wage) would be characterized by the following problem:

$$\begin{aligned} V\left(K_t, Z_t\right) \;=\; & \max_{I_t, L_t}\{F\left(K_t, L_t, Z_t\right) - w\left(Z_t\right) L_t - I_t - \kappa\left(I_t, K_t, Z_t\right) \\ & + \mathbb{E}_t\left[\mathcal{M}_{t+1} V\left(\left(1-\delta\right) K_t + I_t, Z_{t+1}\right)\right]\} \end{aligned} \tag{65}$$

Hence, we can trace how an inflow into equities increases equity prices, lowers the risk premium, and increases investment. We leave the full, quantitative analysis of this to future research, but hope that this will help economists see more concretely how all fits together.

5.5 Calibration of the general equilibrium model

We now calibrate the model. We use the parameter values given in Table 4, which are all presented in annualized terms for clarity. We provide a summary discussion of our parameter choices here, leaving some details to Section E.13. Risk aversion is moderate, at $\gamma = 2$. The macroeconomic parameter values are standard, except for the pure rate of time preference.[64] The parameters concerning the representative fund's mandate, θ, κ, and the activity parameter m_p, were all discussed in Section 3.4. We set a speed of mean reversion of the behavioral disturbance of $\phi_b = 4\%$/year, which induces the same speed of mean reversion for flows $\tilde{f}_t$ and for the P/D ratio. Likewise, we choose its standard deviation to generate the requisite volatility of flows. For parsimony, we assume zero correlation between flow shocks and dividend shocks.

[64] The pure rate of time preference is the only non-standard value: to get a small risk-free rate of 1%, we need to make the agents very patient, so that $\beta > 1$. Indeed, this comes from the Ramsey equation (62), which is $r_f \simeq -\ln\beta + \gamma g$ (neglecting precautionary effects, which are very small in our calibration) with $\gamma g = 4\%$. It would be easy to amend that, for example by adding a small probability of a disaster risk *in flows* (that is, a small probability of a very large negative outflow, which would result in crash risk) or by using Epstein-Zin preferences. We do not do that, because we do not wish to complicate the model. If we normalized the average growth rate to zero, we would not have this difficulty.

Table 4: Parameter values used in the calibration

Variable	Value
Growth rate of endowment and dividend	$g = 2\%$
Std. dev. of endowment growth	$\sigma_y = 0.8\%$
Std. dev. of dividend growth	$\sigma_D = 5\%$
Mixed fund's equity share	$\theta = 0.85$
Mixed fund's sensitivity to risk premium	$\kappa = 1$
Active fraction of funds	$m_p = 0.84$
Mean reversion rate of behavioral disturbance	$\phi_b = 4\%$
Std. dev. of innovations to behavioral disturbance	$\sigma_b = 3.3\%$
Time preference	$\beta = 1.03$
Risk aversion	$\gamma = 2$

Table 5: Moments generated by the calibration

Variable	Value
Macro elasticity	$\zeta = 0.16$
Macro elasticity with mean-reverting flow	$\zeta^M = 0.2$
Macro market effective discount factor, $\rho = \zeta/\kappa$	$\rho = 16\%$
Risk free rate	$r_f = 1\%$
Average equity premium	$\bar{\pi} = 4.4\%$
Average dividend-price ratio	$\delta = 3.4\%$
Std. dev. of stock returns	$\sigma_r = 15\%$
Share of variance of stock returns due to flows	89%
Share of variance of stock returns due to fundamentals	11%
Mean reversion rate of cumulative flow and $\log D/P$	$\phi_f = 4\%$
Std. dev. of innovation to cumulative flow	$\sigma_{\tilde{f}} = 2.8\%$
Slope of log price deviation to flow	$b_f^p = 5$
Slope of equity premium to flow	$b_f^\pi = -0.37$

Table 6: Some stock market moments and predictive regressions

(a) Stock market moments

	Data	Model
Std. dev. of excess stock returns	0.17	0.15
Mean P/D	37	33
Std. dev. of $\log P/D$	0.42	0.5

(b) Predictive regressions

	Data			Model			
Horizon	Slope	S.E.	R^2	Mean of slope	95% CI of slope	S.E.	R^2
1 yr	0.11	(0.034)	0.07	0.14	$[0.04, 0.32]$	(0.048)	0.09
4 yr	0.36	(0.14)	0.18	0.61	$[0.18, 1.19]$	(0.17)	0.28
8 yr	1.00	(0.34)	0.40	1.34	$[0.39, 2.50]$	(0.31)	0.43

Notes. The data are for the United States for 1947-2018, and are calculated based on the CRSP value-weighted index. The predictive regressions for the expected stock return in panel (b) are $R_{t\to t+T} = \alpha_T + \beta_T \log\left(\frac{D_t}{P_t}\right)$, at horizon T (annual frequency). S.E. denotes the Newey-West standard errors with 8 lags. 95% CI denotes the 95% confidence interval of the estimated coefficients on the simulated data. Each run in the simulation uses 72 years.

Table 5 shows the resulting moments implied by the model. It verifies that we match all the "classic" moments, for instance the risk free rate, the average equity premium, and the volatility of stock returns. We see that the model features a large "excess volatility": the flow shocks account for almost 90% of the variance of stock returns. It may be surprising that we can match the equity premium without any of the "modern" asset pricing ingredients, such as a very high risk aversion or disaster risk. The reason is that the preferences of our behavioral investors feature "narrow framing", which leads to an average risk premium given by $\bar{\pi} = \gamma\sigma_r^2$.

Table 6 shows more moments specific to the stock market. We broadly match the volatility of the log P/D ratio, its speed of mean reversion, and the predictive power of forecasting regressions with that P/D ratio.

We conclude that our general equilibrium model with "inelastic markets" is competitive with other widely-used general equilibrium models that match equity market moments. Its main advantage, as we see it, is that is relies on an observable force, flows in and out of equities. In addition, it matches our evidence on the macro elasticity of the market. Also, it retains the CRRA structure, so it is easier to mesh with the basic macro models. Hence, it might be a useful prototype demonstrating how to think about inelastic market in general equilibrium.

6 Finance with Inelastic Markets: Government Policy, Corporate Finance, and Endogenous Flows

We now examine how a number of issues in finance change when markets are inelastic: government and corporate policies, pricing kernels, the modeling of different asset classes, and endogenous flows.

6.1 Governments might stabilize the stock market via quantitative easing in equities

In inelastic markets, the government might prop up asset values, perhaps in times of crisis, or to help firms invest by raising equity at a high price. Indeed, suppose that the government buys f^G percent of the market, and keeps it forever. Then, the market's valuation increases by $p = \frac{f^G}{\zeta} \simeq 5f^G$.[65] So, if the government buys 1% of the market (which may represent roughly 1% of GDP), the market goes up by 5%.[66]

This is what a number of central banks have done. In August 1998, the Hong Kong government, under speculative attack, bought 6% of the Hong Kong stock market: this resulted in a 24% abnormal return, which was not reversed in the following eight weeks (Bhanot and Kadapakkam (2006)). This effect is not entirely well-identified, but is consistent with a large price impact multiplier $\frac{1}{\zeta}$, around 4. Likewise, the Bank of Japan owned 5% of the Japanese stock market in March 2018 (Charoenwong et al. (2020)) and the Chinese "national team" (a government outfit) owned a similar 5% of Chinese stocks in early 2020.[67] In inelastic markets, this may have a large price impact.[68] Those government purchases of equities offer a potentially attractive government policy, as they increase market values, hence lower the cost of capital for firms, and relax credit constraints. So, they might increase hiring and real investments by firms, and GDP. This seems like an interesting topic for research. Brunnermeier et al. (2020) caution about potentially adverse effect if the government's purchases might become too central.

6.2 Corporate finance in inelastic markets

Imagine that firms (the aggregate corporate sector) buy back share in one period, reducing dividends and hence keeping total payouts constant. What happens?

In a frictionless model, this does not affect the firm's value, as per Modigliani-Miller.In an inelastic model, it should now be clear that buybacks will increase the aggregate value of equities. How much depends on the rationality of households, as we now detail. For clarity and brevity, we focus on the two-period model (the same economics holds with an infinite horizon, but the expressions are more complicated; see Section E.7). At time 0, we imagine the representative firm

[65] Note that we assume that investors don't change their holdings to counteract the government's holdings, meaning that Ricardian equivalence does not hold, perhaps because of a form on inattention to the government's actions (Gabaix (2020)).

[66] If the government buys it for just T periods, the impact is $p = \left(1 - \frac{1}{(1+\rho)^T}\right) \frac{f^G}{\zeta}$. Set $f_t = f^G 1_{0 \leq t < T}$ in (16). With the above calibration, this can be a moderate dampening if T is large enough.

[67] Lockett, Hudson. "How the invisible hand of the state works in Chinese stocks." *Financial Times*, 2/4/2020.

[68] We are not aware of a quantification of the macro elasticity for Japan. Barbon and Gianinazzi (2019) and Charoenwong et al. (2020) quantity a micro elasticity – the differential impact on individual stocks that are owned vs. not owned by the government.

buys back a fraction b of the equity shares, where b is small (so that the new number of shares is $Q_0' = Q_0(1-b)$). The buyback is financed by a fall in the time-0 dividend, so the total dividend payout falls from $\mathcal{D}_0$ to $\mathcal{D}_0' = \mathcal{D}_0 - P_0 Q_0 b$, where P_0 is the ex-dividend price, and $P_0 Q_0 b$ is used to finance the share buyback.

We need to take a stand on the households' reaction to those buybacks. Call μ^D (respectively μ^G) the fraction of the change in dividends (respectively, of the change in capital gain) that is "absorbed" by the households – that is, consumed or reinvested in the pure bond fund. If the extra dividend (respectively extra capital gain) is X dollars, consumers will "remove from the mixed fund" $\mu^D X$ (respectively $\mu^G X$) dollars. As households' marginal propensity to consume is higher after a \$1 dividend rather than a \$1 capital gain (Baker et al. (2007)), it is likely that $0 < \mu^G < \mu^D < 1$. We do not seek here to endogenize μ^D and μ^G, which would be a good application of limited attention. We simply trace their implications for the price impact of share buybacks in the following proposition (which is proved in Section E.1).

Proposition 8. (Impact of share buybacks in a two-period model) *Suppose that at time 0 corporations buy back a fraction b of shares, lowering their dividend payments by the corresponding dollar amount, hence keeping total payout constant at time 0. Then, the aggregate value of equities moves by a fraction*

$$v = \frac{\left(\mu^D - \mu^G\right)\theta}{\zeta + \mu^G \theta} b, \tag{66}$$

where μ^D (respectively μ^G) is the fraction of the change in dividends (respectively change in capital gains) "absorbed" by households, i.e. removed from the mixed fund. If $\mu^D > \mu^G$ (so that the marginal propensity to consume out of dividends is higher than that out of capital gains), then share buybacks increase the aggregate market value: $v > 0$.

A very tentative calibration Using the recent estimates of Majlesi et al. (2020), we set $\mu^D \simeq 0.5$ and $\mu^G \simeq 0.03$.[69] Then, (66) says that a buyback of 1% of the market increases the market capitalization by 1.8%. The generalization to an infinite horizon (Section E.7.1) confirms this calibration.

6.3 When flows respond to risk premia in the long run

The present paper mostly points out how impactful flows are. We provided one microfoundation for flows in the macro model of Section 5.1, via the "behavioral disturbance" b_t, which is a stand-in for the forces driving flows. Here, we examine variants of that formulation.

A necessary trend and cycle decomposition for flows First, we record that all models of flows should satisfy the following decomposition to keep the price-dividend ratio stationary.

[69] The above literature does not exactly measure μ^D and μ^G: it measures the impact on consumption, not on consumption plus reallocation to pure bond funds. It is conceivable that some of the capital gains or dividends are reinvested in bonds, even if they're not consumed. So, μ^D (resp. μ^G) is likely to be higher than the marginal propensity to consume out of dividends (resp. capital gains). In addition, what matters is the "long run" propensity, which is hard to measure. One upshot is that it would be interesting for the empirical literature to estimate μ^D, μ^G.

Lemma 1. *(Trend-cycle decomposition for flows) The price-dividend ratio is stationary if and only if the cumulative flow f_t admits the decomposition*

$$f_t = (1-\theta)\, d_t + \hat{f}_t, \tag{67}$$

where d_t the realized long term deviation in dividends ($D_t = D_0 e^{d_t}$), which is typically nonstationary, θ is the equity-weighted equity share, and $\hat{f}_t$ is stationary.

Proof. Recall (75), $q^D = f_t - (1-\theta)\, p_t + \kappa \hat{\pi}_t$. As $q^D = 0$, (67) holds, with $\hat{f}_t = -\kappa \hat{\pi}_t + (1-\theta)\,(p_t - d_t)$. □

How does the market equilibrate in the long run? One might ask, how does the market discover the trend $(1-\theta)\, d_t$ in (67)? It comes out in the model of Section 5.1, via the assumption of (partial) rational expectations. But what about other models? It turns out that a variety of plausible models of investor behavior also lead to stationarity. We briefly summarize the situation, while Section E.6 provides proof and complements.Consider a behavioral rule of the type

$$\Delta f_t = \chi \hat{\pi}_t + \varepsilon_t, \tag{68}$$

with $\chi > 0$: this means that people invest more in equities when they are undervalued, which makes flows stabilizing. Then, one can show that this leads to a stationary P/D ratio, as in Lemma 1, and hence the correct representation (67).[70]

This rule, in turn, generates the following realistic dynamics. We provide the expression in the limit of small time intervals, as the expressions are simpler, and in the case $d_t = 0$ to simplify the analysis.

Proposition 9. (Equilibrium when flows respond to the equity premium in a noisy fashion) *In the limit of small time intervals, the specification of flows* (68) *with i.i.d. shocks ε_t generates a deviation of the price from trend equal to:*

$$\mathbb{E}_0 p_t = \frac{1}{\zeta + \kappa\phi} \mathbb{E}_0 f_t, \qquad \mathbb{E}_0 f_t = (1-\phi)^t\, f_0. \tag{70}$$

The speed of mean-reversion ϕ is the positive solution of $\kappa\phi^2 + (\zeta - \chi)\,\phi = \chi\delta$. The speed of mean-reversion ϕ is increasing in the intensity of the response to the equity premium, χ, and decreasing in ζ and κ. It is zero if $\chi = 0$.

For instance, consider a flow shock f_0 at time 0. Then, the dynamics are those in (20). This captures that endogenously, flows are "digested" by the market at a rate ϕ, which is higher when χ

[70] Rule (68) can have microfoundations of the "behavioral inattention" type:

$$f_t = m f_t^r + (1-m)\, f_{t-1} + \tilde{f}_t, \tag{69}$$

where f_t^r is the rational flow for an investor embedded in this economy, and f_{t-1} is the "default behavioral flow", corresponding to no action, $\tilde{f}_t$ is a stationary "behavioral disturbance", and $m \in [0,1)$ (along with the size of $\tilde{f}_t$) smoothly parametrizes the degree of rationality of the model. This captures that agents are "partially rational", but are also affected by some disturbance $\tilde{f}_t$. Because the rational flow (maximizing $\mathbb{E}_t\left[V^p\left(R_{t+1}\right)\right]$) is $f_t^r = f_t + \frac{\hat{\pi}_t}{\bar{\pi}}$, behavior (69) generates (68) with $\chi = \frac{m}{1-m}\frac{1}{\bar{\pi}} > 0$ and $\varepsilon_t = \frac{\tilde{f}_t}{1-m}$. Formulation (69) extends more easily than (68) to other contexts (Gabaix (2014)).

is higher, i.e. when investors chase risk premia more aggressively (see Bouchaud et al. (2009) for a survey, more geared towards shorter time scales).

Illustrative calibration. If we assume $\chi = 0.23$, we replicate a slow mean-reversion of the P/D ratio of $\phi_f \simeq 4\%$ per year.[71]

One could imagine agents with other behavioral rules, or agents optimizing on the parameters χ, m, this way providing additional cross-asset predictions.[72] We leave that to future research.

7 Conclusion

This paper finds, both theoretically and empirically, that the aggregate stock market is surprisingly price-inelastic, so that flows in and out of the market have a significant impact on prices and risk premia. We refer to this as the inelastic markets hypothesis. We provide tools to analyze inelastic markets, with a simple model featuring key elasticities and an identification strategy using the recently developed method of granular instrumental variables, conceived for this project and laid out in detail in Gabaix and Koijen (2020).

We emphasize though that the "inelastic market hypothesis" remains just that: a hypothesis. Our empirical analysis relies on a new empirical methodology and on fairly unexplored data in this context. An important takeaway from this paper is that the demand elasticity of the aggregate stock market is a key parameter of interest in asset pricing and macro-finance, just like investors' risk aversion, their elasticity of inter-temporal substitution, and the micro elasticity of demand. We provide a first estimate, and we hope that future research will explore other identification strategies to improve and sharpen this estimate.

If the inelastic market hypothesis is correct, it invalidates or qualifies a number of common views in finance and it provides new directions to answer longstanding questions in finance. We outline and then discuss those tenets.

How tenets of finance change if the inelastic markets hypothesis is correct

"Permanent price impact must reflect information." In Proposition 5, a one-time, non mean-reverting inflow permanently changes prices (as in $p = \frac{f}{\zeta}$), even if it contains no information whatsoever. This is plainly because a permanent change in the demand for equities must permanently change their equilibrium price – and this effect is quantitatively important in inelastic markets. The typical empirical strategy to look for reversals as signs of flows (rather than information) moving prices does not work in this case. By the same logic, we can see large changes in prices but small changes in long-horizon expected returns.

"Fast and smart investors (perhaps hedge funds) will provide elasticity to the market." This is not true: in part because hedge funds are small (they own only about 5% of the market, see Section 2.2), they cannot provide much elasticity for the market as a whole (so ζ remains low), even though they might ensure short term news are incorporated quickly (so that κ is quite high). In addition,

[71] The parameter χ is unitless: in continuous time, $df_t = \chi \hat{\pi}_t dt + \sigma dz_t$.

[72] Alternatively, consider a rule like:

$$\Delta f_t = \chi \hat{\pi}_t + \beta (d_t - p_t) + \Delta \tilde{f}_t, \tag{71}$$

where χ and β are weakly positive, one of them is strictly positive, and $\tilde{f}_t$ is an AR(1). The coefficients χ and β are "stabilizing" forces: they make investors buy when expected returns are high. Then, the rule (71) also leads to the correct form shown in Lemma 1. However, a rule like $f_t = \chi \hat{\pi}_t + \beta (d_t - p_t) + \tilde{f}_t$ would not lead to a stationary P/D ratio: while the right-hand side would be stationary, by Lemma 1 the left-hand side should not be stationary.

those smart-money investors often face risk constraints that limit their ability to aggressively step in during aggregate downturns.

"Trading volume is very high, so the equity market must be very elastic." Trading volume in the equity market is high (about 100% of the value of the market each year), but most of it exchanges one share for another share (perhaps via a round-trip through cash). These trades within the universe of equities do no count toward the aggregate flow f. As we mentioned above, the sum of the absolute values of sector-level flows, including net issuances, relative to the size of the market, averages only 1.9% per year.

"For every buyer there is a seller; so, saying 'there was an increase in the demand for equities' is meaningless". Economists often appeal to the truism that "for every buyer there is a seller" to disregard the notion that a measurable increase in the willingness of the average trader to buy more of the market will push prices up ("buying pressure"). Our model clarifies that this reasoning is incorrect. In Proposition 2, f is the pressure to buy stocks (if it is positive), and the demand $q^D = -\zeta p + f$ has a component $-\zeta p$ expressing that "sellers" will sell the shares to "buyers" represented by f. So there are both buyers and sellers (or really, a force making the representative fund buy, and a force making it sell), but at the same time, buying pressure f does move the price by $p = \frac{f}{\zeta}$. Moreover, it is directly measurable via the change in asset holdings (bonds in the case of the undergraduate example of Section 3.1).

"The market often looks impressively efficient in the short run, so it must be quite macro-efficient." The contrast between the market's "short run efficiency" and "macro-efficiency" is sharp in equation (16): future events are discounted at a rate $\rho = \frac{\zeta}{\kappa} = \delta + \frac{1-\theta}{\kappa}$, so that a highly far-sighted market has a lower value of ρ. So, the market can be very forward looking (low ρ), even if it is very macro-inelastic (low ζ), provided that "far-sightedness" κ is relatively high compared to ζ (for example, because there are a few powerfully forward-looking arbitrageurs). As an example, consider the announcement of an event that will take effect in a week, such as a permanent increase in dividends or inflows. In our calibration, the market's current reaction to the announcement is a fraction 99.8% of the eventual present value of the future dividends or inflows.[73] In that sense, the market looks impressively efficient. But again, it is "short-term predictability efficient" (it smooths announcements) and "micro efficient" (it processes well the relative valuations of stocks), but it is not "macro efficient" (as Samuelson (1998) put it) or "long-term predictability efficient" – it does not absorb well very persistent shocks. Furthermore, even though prices respond promptly around major events, it is generally hard to assess whether the market moved by just the right amount, or instead under- or over-reacted. In addition to a large literature demonstrating drifts in prices before and after macro events (such as Federal Open Market Committee meetings), our model implies that persistent flows around such events can lead to persistent deviations in prices, and typical event study graphs that do not display much of a drift in prices following the event would be uninformative about macro efficiency.

"Share buybacks do not affect equity returns, as proved by the Modigliani-Miller theorem." In the traditional frictionless model, the return impact of a share buyback should be zero. However, in our model, if firms in the aggregate buy back \$1 worth of equity, that increases aggregate valuations (Section 6.2 detailed this). Hence, share buybacks are potentially a large source of fluctuations in the market. In our model, a combination of fund mandates and consumers' bounded rationality leads to a violation of the Modigliani-Miller neutrality. More broadly, corporate actions such as share issuances, transactions by insiders, et cetera, may have a large impact on prices beyond any

[73] Indeed, $(1+\rho)^{-T} = 99.8\%$, taking the ρ calibrated in Section 3.4 and $T = 1/52$ years.

informational channel. Most extant empirical evidence focuses on announcements at the firm level, while we emphasize their impact in aggregate. By focusing on well-identified firm-level responses, one identifies the micro-elasticity, not the macro elasticity ζ. It will be interesting to explore in detail how important corporate decisions are for fluctuations in the aggregate stock market.

We next discuss a few questions that seem important for future research.

Why is the aggregate demand for equities so inelastic? The core of the inelastic markets hypothesis is that the macro demand elasticity ζ is low. Why is it so low? We highlighted two reasons, namely fixed-share mandates, such as those of many funds that are 100% in equities and hence have zero elasticity (and in general $\zeta > 0$, $\kappa = 0$), and inertia (i.e., some funds or people are just buy-and-hold, creating $\zeta = \kappa = 0$). This may be due to a taste for simplicity, or to agency frictions: as the household is not sure about the quality of the manager, a simple scheme like a constant share in equities may be sensible – otherwise the manager may take foolish risks. There are other possibilities. If some funds have a Value-at-Risk constraint, and volatility goes up a lot in bad times, they need to sell when the markets fall, so that their ζ and κ are negative. A different possibility is that when prices move, people's subjective perception of the equity premium does not move much. One reason might be that investors think the rest of the market is well-informed. Also, going from market prices to the equity premium is a statistically error-prone procedure, so that market participants shrink towards no reaction to this (Black (1986), Summers (1986)). Alternatively, many investors may not place much weight on the price-earnings ratio as a reliable forecasting tool, perhaps because they want parsimonious models and price-dividend ratios are not that useful as short-run forecasters, or because many investors just do not wish to bother paying attention to them (Gabaix (2014), Chinco and Fos (2019)). The pass-through between subjective beliefs and actions might be low, as it is for retail investors (Giglio et al. (2019)). Finally, demand may respond little to prices because demand shocks are highly persistent.[74] In the end, while pinpointing the exact reasons for low market elasticity would be quite interesting, this question has a large number of plausible answers. Fortunately, it is possible to write a framework in a way that is relatively independent to the exact source of low elasticity, and this is the path we chose.

What are the determinants of flows? It is clear that it would be desirable to know more about the determinants of flows at a high frequency. We provided a minimalist model with a "behavioral disturbance" (which was enough to study its general equilibrium impact), and some simple correlations in Section 4.5, but this is clearly a first pass. Establishing the various channels of flows could be a whole line of enquiry, perhaps with micro data such as those used by Calvet et al. (2009) or Giglio et al. (2019). To appreciate the richness of those determinants, let us observe that flow shocks could come from varied sources, such as: (i) changes in beliefs about future flows or fundamentals, as these both affect expected returns, per Proposition 5; (ii) "liquidity needs", for instance insurance companies selling stocks after a hurricane; (iii) more generally, heterogeneous

[74] For instance, imagine a very simple model $f_t = \sum_k F_k \mathbb{I}\left(t \in \left[\tau_k^0, \tau_k^1\right]\right)$, where F_k is constant, $\left[\tau_k^0, \tau_k^1\right]$ the period of time that a flows stays in the market, and $\tau_k^1 - \tau_k^0 \sim \exp(\lambda)$. From an institutional perspective, one can also imagine that a large asset manager launches a fund that attracts capital, and that this capital is sticky, but the period for which it stays is unclear. If λ is low, then prices will respond sharply to the flow, even though the expected return does not move much. Uncertainty about the persistence of the demand shock introduces uncertainty about how the price change maps to expected returns, leading to a muted response and a low ζ. This model is in quite sharp contrast with the traditional view in which flows have a temporary price impact (for instance Coval and Stafford (2007)).

income or wealth shocks to different groups (including foreign versus domestic investors) changing the effective propensity to invest in stocks by the average investor; (iv) corporate actions by firms such as decisions to buy back or issue shares; (v) shocks to substitute assets, which might for example prompt investors to rebalance towards stocks when bond yields go down; (vi) changes in the advertising or advice by institutional advisers, as explored in Ben-David et al. (2020b); (vii) "road shows" in which firms or governments try to convince potential investors to buy into a prospective equity offering or privatization; (viii) mechanical forced trading via so-called "delta hedging," whereby traders who have sold put options and continuously hedge them need to sell stocks when stock prices fall.

Some outstanding questions The inelastic markets hypothesis, in the framework we provide, makes a number of issues that were boring, irrelevant, or inaccessible suddenly interesting, important, and researchable, thanks to the theoretical and empirical framework developed in this paper. What created flows in the major episodes (so that one could write a flow-based history, explaining episodes such as that of Figure 2)? Do they impact investment? How much can and should governments intervene in equity markets? Do share buybacks account for a large share of activity? What determines the low elasticity ζ of markets, and how forward-looking are the policies of funds (κ)? Generalizing, what are the cross-market elasticities, meaning the forces that create "contagion" across market? Also, these same effects will generalize to other markets (such as the markets for corporate bonds and currencies): if so, how and what are the policy implications? This is a rich number of questions that hopefully economists will be able to answer in the coming years.

A Appendix: Proofs

Complement to the Proof of Proposition 2 *Next, consider the case with a general κ.* Taking logs and then deviations from the baseline D/P ratio gives:

$$\Delta \ln \frac{D^e}{P} = \Delta \ln D^e - \Delta \ln P = d - p$$

On the other hand, as $\delta = \frac{D^e}{P} = 1 + r + \pi$, we have $\Delta \ln \frac{D^e}{P} = \frac{\Delta \pi}{1+r+\pi} = \delta \hat{\pi}$ (with $\hat{\pi} = \Delta \pi$), so that

$$\hat{\pi} = \delta \left(d - p \right). \tag{72}$$

Using (9),[75]

$$q^D = w - p + \kappa \hat{\pi}, \tag{73}$$

that is

$$q^D = -\left(1 - \theta\right) p + f + \kappa \delta \left(d - p \right) = -\left(1 - \theta + \kappa \delta\right) p + f + \kappa \delta d,$$

which yields (7).

[75] We take logs in (3), so that $\ln Q^D = \ln W + \ln \theta - \ln P + \kappa \hat{\pi}$. Given that initially $\ln Q^D = \ln \bar{W} + \ln \theta - \ln \bar{P}$, taking differences we have

$$\Delta \ln Q^D = \Delta \ln W - \Delta \ln P + \kappa \hat{\pi}.$$

Finally, we use the Taylor expansion $\Delta \ln W \simeq w$ and $\Delta \ln P \simeq p$ to yield $q^D \simeq w - p + \kappa \hat{\pi}$.

Proof of Proposition 4 The wealth deviation is, using the same derivation as in the two-period model, $w_t = \theta p_t + f_t$, so that the demand response is given by the linearization of (12):

$$\begin{aligned} q_t^D &= w_t - p_t + (\kappa\hat{\pi}_t + \nu_t) \\ &= (f_t + \theta p_t) - p_t + (\kappa\hat{\pi}_t + \nu_t) \quad (74) \\ &= f_t + \nu_t - (1-\theta)\, p_t + \kappa\left(\mathbb{E}_t \Delta p_{t+1} + \delta\left(d_t^e - p_t\right)\right), \quad (75) \end{aligned}$$

which yields (15).

Proof of Proposition 5 Equation (15) can be rewritten as $q^D = \kappa\left(\mathbb{E}_t \Delta p_{t+1} - \rho p_t + \delta d_t^e\right) + f_t + \nu_t$. As $q^D = 0$, this is also:

$$\mathbb{E}_t \Delta p_{t+1} - \rho p_t + \delta d_t^e + \frac{\nu_t + f_t}{\kappa} = 0. \quad (76)$$

Defining $z_t \coloneqq \delta d_t^e + \frac{\nu_t + f_t}{\kappa}$, this gives $p_t = \frac{\mathbb{E}_t p_{t+1} + z_t}{1+\rho}$, so that $p_t = \mathbb{E}_t \sum_{\tau \geq t}^{\infty} \frac{z_\tau}{(1+\rho)^{\tau - t + 1}}$. The equity premium comes from (74).

Proof sketch of Proposition 7 We provide a proof sketch here, leaving the more delicate points to the full derivation in Section E.1. First, as we say, because the consumer part of the household is rational and can trade the riskless bonds, her Euler equation holds, so that (49) holds. This pins down the risk-free rate, which is constant, $R_f = \frac{1}{\beta \mathbb{E}\left[G_{t+1}^{-\gamma}\right]}$.

Next, given all net wealth is in equities (for the representative agent), the total return on wealth is $R_{t+1}^{\mathcal{E}} = \frac{P_{t+1} + Y_{t+1}}{P_t}$. Assumptions (50)-(52) imply that the financier is "rational on average", so that the Euler equation for stocks holds on average. Given that the financier maximizes a "narrow-frame" utility of wealth $V\left(W_{t+1^-}\right)$, and $W_{t+1^-} = W_t R_{t+1}^{\mathcal{E}}$, the first order condition for stock allocation is $\mathbb{E}\left[W_{t+1^-}^{-\gamma}\left(R_{t+1}^{\mathcal{E}} - R_f\right)\right] = 0$, implying $\mathbb{E}\left[\left(R_{t+1}^{\mathcal{E}}\right)^{-\gamma}\left(R_{t+1}^{\mathcal{E}} - R_f\right)\right] = 0$. Taking the limit of small time intervals, this means that the average equity premium is $\bar{\pi} = \gamma\sigma_r^2$. The value of p_t, $\hat{\pi}_t$ and so on, were derived in (22).□

B Appendix: Identification methodology

We summarize the algorithms that we use to estimate the multipliers and elasticities in Section 4.2 and the multipliers in Section 4.3. The algorithms are the same, with some minor adjustments given the unique features of either the FoF data or 13F data.

B.1 Algorithm used for sector-level data

We summarize the algorithm that we use for the Flow of Funds (FoF) data in Section 4.2.

1. For efficiency, we construct pseudo-equal value weights $\tilde{E}_{j,t-1}$, where we start from $\tilde{E}_j^{\sigma} = \frac{\sigma_j^{-2}}{\sum_{k=1}^{N} \sigma_k^{-2}}$,where $\sigma_j = \sigma\left(\Delta q_{jt}\right)$, and define $\tilde{E}_j = \min\left\{\xi \tilde{E}_j^{\sigma}, \frac{1.5}{N}\right\}$, where $\xi \geq 1$ is tuned so that $\sum_j \tilde{E}_j = 1$. This winsorizes the quasi-equal weights to be at most 50% higher than strict

equal weights.[76] This adjustment ensures that the equal weights are not too concentrated for sectors with very stable Δq_{jt}. This is relevant when the number of sectors is small, as is the case for the FoF.

2. We run the panel regression

$$\Delta q_{jt} = \alpha_j + \beta_t + \gamma_j \Delta y_t + \delta_j t + \Delta \check{q}_{jt}, \tag{77}$$

using $\tilde{E}$ as regression weights, and construct the $\Delta \check{q}_{jt}$ as the residuals. Here Δy_t is quarterly real GDP growth and we allow for a time trend as some sectors grew substantially faster in, for instance, the nineties than the subsequent period. We exclude the corporate sector in running the panel regression and in constructing the instrument.

3. We extract the principal components of $\tilde{E}_j^{\frac{1}{2}} \Delta \check{q}_{jt}$ and denote the estimated vector of principal components by $\eta_t^{PC,e}$. In extracting the factors, we once again exclude the corporate sector.

4. We construct the GIV instrument, excluding the corporate sector:[77]

$$Z_t = \sum_{j=1}^{N} S_{j,t-1} \Delta \check{q}_{it}. \tag{78}$$

5. We estimate the multiplier, M, using the time-series regression

$$\Delta p_t = \alpha + M Z_t + \lambda_P' \eta_t^e + e_t, \tag{79}$$

where $\eta_t^e = \left(\Delta y_t, \eta_t^{PC,e}\right)$. This regression is also the first stage to estimate the elasticities. Instrumenting Δp_t by Z_t in both cases, we estimate the demand elasticity via

$$\Delta q_{Et} = \alpha_E - \zeta \Delta p_t + \lambda_E' \eta_t^e + e_t, \tag{80}$$

and the supply elasticity via

$$\Delta q_{Ct} = \alpha_C - \zeta_C \Delta p_t + \lambda_C' \eta_t^e + e_t. \tag{81}$$

B.2 Algorithm used for investor-level data

We summarize the algorithm that we use for the Flow of Funds (FoF) data in Section 4.3.

1. We standardize the characteristics cross-sectionally in each period by removing their means and dividing by their standard deviations.

[76] Quasi-equal weights $\tilde{E}_j$ are preferable to equal weights $E_j = \frac{1}{N}$ as they add precision — in the same way to estimate a mean, weighing by inverse variance is better than equal weighing (Gabaix and Koijen (2020)). The primary objective of the winsorization at $\frac{1.5}{N}$ is to downplay the importance of very volatile sectors that may distort the estimation of the common factors. If the inverse variance weights get too concentrated, the same concern applies to very stable sectors in that case. While 50% is somewhat arbitrary, it is a significant departure from equal weights. We also explore the sensitivity of our results to this cutoff and find them to be robust.

[77] An equivalent way to proceed is to use $z_t = \sum_{j=1}^{N} S_{j,t-1} \check{u}_{jt}$, where $\check{u}_{jt}$ is the measure of idiosyncratic shock common from step 4. This way, z_t is made of idiosyncratic shocks. As we control for $\eta_t^{PC,e}$ below, the two procedures are equivalent.

2. We run the panel regression

$$\Delta q_{jt} = \alpha_j + \beta_t + \gamma_j \Delta y_t + \delta_j I_{t-1} + x'_{j,t-1}\eta_t^x + \Delta\check{q}_{jt},$$

and construct the $\Delta\check{q}_{jt}$ as the residuals. We use $\frac{1}{\sigma_j^2}$, with $\sigma_j = \sigma(\Delta q_{jt})$,[78] as regression weights, Δy_t being quarterly real GDP growth, and I_t denotes institutional ownership to capture the time trend as in (77).[79]

3. We also allow for the possibility of additional common factors, which we estimate using principal components analysis. As the panel is unbalanced, we estimate the factors in

$$\Delta\check{q}_{jt} = \check{\lambda_j}'\eta_t + \check{u}_{jt},$$

for K latent factors, η_t^{PC}, using the objective, with $\Delta\dot{q}_t = \frac{\Delta\check{q}_{jt}}{\sigma_j}$ to adjust for heteroscedasticity,

$$\min_{(\eta_t,\dot{\lambda})} \sum_{t=1}^{T}\sum_{k=1}^{N_t}\left(\Delta\dot{q}_t - \dot{\lambda}'_k\eta_t\right)^2,$$

under the normalization $\sqrt{\dot{\lambda}'_k\dot{\lambda}_k} = 1$, for each factor k. We estimate the factors recursively and estimate the residuals, $\check{u}_{jt}$, by regressing $\Delta\check{q}_{jt}$ on the estimated factor (using the same regression weights as before) for each sector. For each of the factors, we estimate the factor loadings, λ_k, and factor realizations, η_t, by repeated cross-sectional and time-series regressions until convergence.

4. We construct the instrument, $Z_t = \sum_{j\in\mathcal{I}} S_{j,t-1}\Delta\check{q}_{jt}$, where $\mathcal{I}$ is the set of institutions used in the estimation (note that this works even if $\sum_{j\in\mathcal{I}} S_{j,t-1} < 1$).

5. We estimate the multiplier via the regression

$$\Delta p_t = \alpha + MZ_t + \lambda'_P\eta_t^e + e_t,$$

where $\eta_t^e = \left(\Delta y_t, \eta_t^{PC,e}\right)$.[80]

References

Adrian, Tobias and Hyun Song Shin, "Procyclical Leverage and Value-at-Risk," *The Review of Financial Studies*, 10 2013, *27* (2), 373–403.

Amiti, Mary and David E Weinstein, "How much do idiosyncratic bank shocks affect investment? Evidence from matched bank-firm loan data," *Journal of Political Economy*, 2018, *126* (2), 525–587.

Bacchetta, Philippe and Eric Van Wincoop, "Infrequent portfolio decisions: A solution to the forward discount puzzle," *American Economic Review*, 2010, *100* (3), 870–904.

[78] In contrast to the algorithm for the FoF, we do not need to winsorize the regression weights as the number of observations in the cross section is sufficiently large.

[79] Without I_t, or a time trend, the first principal component displays a trend, similar to I_t, as some institutions grew rapidly in the beginning of the century. Hence, to avoid that the first principal component reflects this trend, we include I_t, which removes the trend from the principal components.

[80] The estimated parametric factors, $\eta_t^{x,e}$, may depend on prices and we therefore do not use them as controls.

Baker, Malcolm and Jeffrey Wurgler, "A catering theory of dividends," *The Journal of Finance*, 2004, *59* (3), 1125–1165.

Baker, Malcolm, Stefan Nagel, and Jeffrey Wurgler, "The Effect of Dividends on Consumption," *Brookings Papers on Economic Activity*, 2007, *38* (1), 231–292.

Bansal, Ravi and Amir Yaron, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 2004, *59* (4), 1481–1509.

Barberis, Nicholas and Andrei Shleifer, "Style investing," *Journal of financial Economics*, 2003, *68* (2), 161–199.

Barbon, Andrea and Virginia Gianinazzi, "Quantitative Easing and Equity Prices: Evidence from the ETF Program of the Bank of Japan," *The Review of Asset Pricing Studies*, 2019, *9* (2), 210–255.

Barro, Robert J, "Rare disasters and asset markets in the twentieth century," *The Quarterly Journal of Economics*, 2006, *121* (3), 823–866.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, "Hedge fund stock trading in the financial crisis of 2007–2009," *The Review of Financial Studies*, 2012, *25* (1), 1–54.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, "Non-Fundamental Demand and Style Returns," 2020. Working paper, Ohio State University.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, "Style Investing, Positive Feedback Loops, and Asset Pricing Factors," *Working Paper*, 2020.

Bhanot, Karan and Palani-Rajan Kadapakkam, "Anatomy of a government intervention in index stocks: Price pressure or information effects?," *The Journal of Business*, 2006, *79* (2), 963–986.

Black, Fischer, "Noise," *The Journal of Finance*, 1986, *41* (3), 528–543.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, "Expectations of Fundamentals and Stock Market Puzzles," Technical Report, National Bureau of Economic Research 2020.

Bouchaud, Jean-Philippe, J Doyne Farmer, and Fabrizio Lillo, "How markets slowly digest changes in supply and demand," in "Handbook of financial markets: dynamics and evolution," Elsevier, 2009, pp. 57–160.

Brainard, William C. and James Tobin, "Pitfalls in Financial Model Building," *American Economic Review: Papers and Proceedings*, 1968, *58* (2), 99–122.

Brunnermeier, Markus K. and Stefan Nagel, "Hedge Funds and the Technology Bubble," *The Journal of Finance*, 2004, *59*, 2013–2040.

Brunnermeier, Markus K, Michael Sockin, and Wei Xiong, "China's model of managing the financial system," Technical Report, National Bureau of Economic Research 2020.

Caballero, Ricardo J and Alp Simsek, "A risk-centric model of demand recessions and speculation," *The Quarterly Journal of Economics*, 2019.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini, "Fight Or Flight? Portfolio Rebalancing by Individual Investors," *Quarterly Journal of Economics*, 2009, *124* (1), 301–348.

Camanho, Nelson, Harald Hau, and Hélène Rey, "Global portfolio rebalancing and exchange rates," *Working Paper*, 2019.

Campbell, John Y. and John H. Cochrane, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 1999, *107* (2), 205–251.

Carvalho, Vasco M and Basile Grassi, "Large firm dynamics and the business cycle," *American Economic Review*, 2019, *109* (4), 1375–1425.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, "Regression Discontinuity and the Price Effects of Stock Market Indexing," *Review of Financial Studies*, 2014, *28* (1), 212–246.

Charoenwong, Ben, Randall Morck, and Yupana Wiwattanakantang, "Bank of Japan Equity Purchases: The Final Frontier in Extreme Quantitative Easing," April 2020. NBER Working Paper 25525.

Chien, YiLi, Harold Cole, and Hanno Lustig, "Is the volatility of the market price of risk due to intermittent portfolio rebalancing?," *American Economic Review*, 2012, *102* (6), 2859–96.

Chinco, Alexander and Vyacheslav Fos, "The sound of many funds rebalancing," 2019.

Cochrane, John H., "Presidential Address: Discount Rates," *The Journal of Finance*, 2011, *66* (4), 1047–1108.

Coimbra, Nuno and Helene Rey, "Financial Cycles with Heterogeneous Intermediaries," 2020. Working Paper.

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

Cutler, David M, James M Poterba, and Lawrence H Summers, "Speculative Dynamics And The Role Of Feedback Traders," *The American Economic Review Papers and Proceedings*, 1990, *80* (2), 63–68.

Da, Zhi, Borja Larrain, Clemens Sialm, and Jose Tessada, "Destabilizing Financial Advice: Evidence from Pension Fund Reallocations," *Review of Financial Studies*, 2018, *31* (10), 3720–3755.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, 1990, *98* (4), 703–738.

Deuskar, Prachi and Timothy C. Johnson, "Market Liquidity and Flow-driven Risk," *The Review of Financial Studies*, 2011, *24* (3), 721–752.

Di Giovanni, Julian and Andrei A Levchenko, "Country size, international trade, and aggregate fluctuations in granular economies," *Journal of Political Economy*, 2012, *120* (6), 1083–1132.

Duffie, Darrell, "Presidential Address: Asset Price Dynamics with Slow-Moving Capital," *Journal of Finance*, 2010, *65* (4), 1237–1267.

Duffie, Darrell and Bruno Strulovici, "Capital mobility and asset pricing," *Econometrica*, 2012, *80* (6), 2469–2509.

Farmer, Roger, *Expectations, employment and prices*, Oxford University Press, 2010.

Frazzini, Andrea and Owen A Lamont, "Dumb money: Mutual fund flows and the cross-section of stock returns," *Journal of financial economics*, 2008, *88* (2), 299–322.

Frazzini, Andrea, Ronen Israel, and Tobias J Moskowitz, "Trading costs," *Available at SSRN 3229719*, 2018.

Friedman, Benjamin M., "Financial Flow Variables and the Short-Run Determination of Long-Term Interest Rates," *Journal of Political Economy*, 1977, *85* (4), 661–689.

Gabaix, Xavier, "The granular origins of aggregate fluctuations," *Econometrica*, 2011, *79* (3), 733–772.

Gabaix, Xavier, "Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance," *The Quarterly journal of economics*, 2012, *127* (2), 645–700.

Gabaix, Xavier, "A sparsity-based model of bounded rationality," *The Quarterly Journal of Economics*, 2014, *129* (4), 1661–1710.

Gabaix, Xavier, "Behavioral Inattention," *Handbook of Behavioral Economics*, 2019, *2*, 261–344.

Gabaix, Xavier, "A behavioral New Keynesian model," *American Economic Review*, 2020, *110* (8), 2271–2327.

Gabaix, Xavier and Matteo Maggiori, "International liquidity and exchange rate dynamics," *The Quarterly Journal of Economics*, 2015, *130* (3), 1369–1420.

Gabaix, Xavier and Ralph SJ Koijen, "Granular instrumental variables," *Available at SSRN 3368612*, 2020.

Gabaix, Xavier, Arvind Krishnamurthy, and Olivier Vigneron, "Limits of arbitrage: theory and evidence from the mortgage-backed securities market," *The Journal of Finance*, 2007, *62* (2), 557–595.

Galaasen, S, R Jamilov, R Juelsrud, and H Rey, "Granular credit risk," Technical Report, Working paper 2020.

Garleanu, Nicolae, Lasse Heje Pedersen, and Allen M. Poteshman, "Demand-Based Option Pricing," *Review of Financial Studies*, 2009, *22* (10), 4259–4299.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, "Five Facts About Beliefs and Portfolios," Working Paper 25744, National Bureau of Economic Research April 2019.

Gourinchas, Pierre-Olivier, Walker Ray, and Dimitri Vayanos, "A preferred-habitat model of term premia and currency risk," *Working Paper*, 2020.

Greenwood, Robin and Andrei Shleifer, "Expectations of returns and expected returns," *The Review of Financial Studies*, 2014, *27* (3), 714–746.

Greenwood, Robin and Dimitri Vayanos, "Bond Supply and Excess Bond Returns," *The Review of Financial Studies*, 2014, *27* (3), 663–713.

Greenwood, Robin and Samuel G Hanson, "Issuer quality and corporate bond returns," *The Review of Financial Studies*, 2013, *26* (6), 1483–1525.

Greenwood, Robin, Samuel G Hanson, Jeremy C Stein, and Adi Sunderam, "A quantity-driven theory of term premiums and exchange rates," *Working Paper*, 2019.

Hamilton, James D., "Why You Should Never Use the Hodrick-Prescott Filter," *The Review of Economics and Statistics*, 2018, *100* (5), 831–843.

Harris, Lawrence and Eitan Gurel, "Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Existence of Price Pressures," *Journal of Finance*, 1986, *41* (4), 815–829.

He, Zhiguo and Arvind Krishnamurthy, "Intermediary Asset Pricing," *American Economic Review*, 2013, *103* (2), 732–770.

Herskovic, Bernard, Bryan T Kelly, Hanno N Lustig, and Stijn Van Nieuwerburgh, "Firm volatility in granular networks," *Journal of Political Economy*, forthcoming.

Johnson, Timothy C, "Dynamic liquidity in endowment economies," *Journal of Financial Economics*, 2006, *80* (3), 531–562.

Kendall, Maurice G, "Note on bias in the estimation of autocorrelation," *Biometrika*, 1954, *41* (3-4), 403–404.

Koijen, Ralph SJ and Motohiro Yogo, "A demand system approach to asset pricing," *Journal of Political Economy*, 2019, *127* (4), 1475–1515.

Koijen, Ralph SJ, Robert J Richmond, and Motohiro Yogo, "Which investors matter for global equity valuations and expected returns?," *Working Paper*, 2019.

Koijen, Ralph SJ, Tobias J Moskowitz, Lasse Heje Pedersen, and Evert B Vrugt, "Carry," *Journal of Financial Economics*, 2018, *127* (2), 197–225.

Kondor, Péter and Dimitri Vayanos, "Liquidity risk and the dynamics of arbitrage capital," *The Journal of Finance*, 2019, *74* (3), 1139–1173.

Kyle, Albert S., "Continuous Auctions and Insider Trading," *Econometrica*, 1985, *53* (6), 1315–1335.

Li, Jennifer, Neil D. Pearson, and Qi Zhang, "Impact of Demand Shocks on the Stock Market: Evidence from Chinese IPOs," 2020. Working paper, Shanghai Advanced Institute of Finance.

Li, Jiacui, "Slow-moving liquidity provision and flow-driven common factors in stock returns," *Available at SSRN 2909960*, 2018.

Lucas, Robert E. Jr., "Asset Prices in an Exchange Economy," *Econometrica*, 1978, *46* (6), 1429–1445.

Ma, Yueran, "Nonfinancial Firms as Cross-Market Arbitrageurs," *The Journal of Finance*, 2019, *74* (6), 3041–3087.

Majlesi, Kaveh, Marco Di Maggio, and Amir Kermani, "Stock Market Returns and Consumption," *Journal of Finance*, 2020.

Martin, Ian, "What is the Expected Return on the Market?," *The Quarterly Journal of Economics*, 2017, *132* (1), 367–433.

Mitchell, Mark, Lasse Heje Pedersen, and Todd Pulvino, "Slow moving capital," *American Economic Review Papers and Proceedings*, 2007, *97* (2), 215–220.

Moreira, Alan, "Capital immobility and the reach for yield," *Journal of Economic Theory*, 2019, *183*, 907–951.

Newey, Whitney K. and Kenneth D. West, "Automatic Lag Selection in Covariance Matrix Estimation," *The Review of Economic Studies*, 10 1994, *61* (4), 631–653.

Rigobon, Roberto, "Identification through Heteroskedasticity," *The Review of Economics and Statistics*, 2003, *85* (4), 777–792.

Samuelson, Paul A, "Summing up on business cycles: opening address," *Conference Proceedings of the Federal Reserve Bank of Boston*, 1998, *42*, 33–36.

Shiller, Robert, "Stock prices and social dynamics," *Brookings papers on economic activity*, 1984, *1984* (2), 457–510.

Shleifer, Andrei, "Do Demand Curves for Stocks Slope Down?," *Journal of Finance*, 1986, *41* (3), 579–590.

Shleifer, Andrei, "Inefficient markets: An introduction to behavioural finance," *Oxford University Press*, 2000.

Summers, Lawrence H, "Does the stock market rationally reflect fundamental values?," *The Journal of Finance*, 1986, *41* (3), 591–601.

Timmer, Yannick, "Cyclical investment behavior across financial institutions," *Journal of Financial Economics*, 2018, *129* (2), 268–286.

Tobin, James, "Monetary policy: recent theory and practice," in "Current Issues in Monetary Economics," Springer, 1998, pp. 13–21.

Vayanos, Dimitri and Jean-Luc Vila, "A preferred-habitat model of the term structure of interest rates," *Working Paper*, 2020.

Vayanos, Dimitri and Paul Woolley, "An Institutional Theory of Momentum and Reversal," *Review of Financial Studies*, 2013, *26* (5), 1087–1145.

Wachter, Jessica A, "Can time-varying risk of rare disasters explain aggregate stock market volatility?," *The Journal of Finance*, 2013, *68* (3), 987–1035.

Warther, Vincent A, "Aggregate mutual fund flows and security returns," *Journal of financial economics*, 1995, *39* (2-3), 209–235.

Wurgler, Jeffrey and Ekaterina Zhuravskaya, "Does Arbitrage Flatten Demand Curves for Stocks?," *Journal of Business*, 2002, *75* (4), 583–608.

Online Appendix for

"In Search of the Origins of Financial Fluctuations: The Inelastic Market Hypothesis"

Xavier Gabaix and Ralph S.J. Koijen

February 4, 2021

Table of contents

C Appendix: Data Sources and Construction **60**
C.1 Data sources . . . 60
C.2 Sector-level data: Flow of Funds . . . 60
C.3 Investor-level data: 13F filings . . . 66
C.4 Mutual fund and ETF data . . . 68
C.5 State and local pension fund data . . . 68

D Additional empirical results **68**
D.1 Performance of the GIV Estimator: Simulations . . . 68
D.2 Drawdown dynamics . . . 69
D.3 Flows across investor classes are small . . . 70
D.4 Estimating κ using fund-level data . . . 71
D.5 GIV estimates using FoF data: Robustness . . . 74
D.6 The volatility of idiosyncratic demand shocks . . . 75
D.7 Screening capital flows . . . 75
D.8 Survey Details . . . 78

E Appendix: Theory Complements **81**
E.1 Appendix: Omitted Proofs . . . 81
E.2 Different asset classes . . . 84
E.3 The model in continuous time . . . 85
E.4 Micro versus macro elasticity: The cross-section of stocks . . . 86
E.5 The model with time-varying market inelasticity . . . 89
E.6 When flows react to the risk premia . . . 89
E.7 Corporate finance in inelastic markets: Complements . . . 94
E.8 General equilibrium in a two-period economy . . . 96
E.9 Details of the household's problem in general equilibrium . . . 98
E.10 Pricing kernel consistent with flow-based pricing: Complements . . . 99
E.11 On the link between the Kyle lambda and the market inelasticity . . . 101
E.12 Short-term versus long-term elasticity when funds are inertial . . . 103
E.13 Calibration of the general equilibrium model of Section 5: Details . . . 106

C Appendix: Data Sources and Construction

C.1 Data sources

Flow of funds We use data from the Flow of Funds (FoF) on equity and bond holdings. To compute bond holdings, we combine the holdings of Treasury bonds and corporate bonds. We discuss the mapping of FoF data to our model in detail below. The sample is quarterly from 1960 to 2018. However, as we explain below, we focus our main estimation results on the sample from 1993.Q1 to 2018.Q4. We use the June 2019 vintage. Data of different vintages can can be downloaded from this website.

Mutual funds We use data on mutual fund flows, assets under management, and the share invested in US equities from Morningstar. The sample is quarterly from 1993.Q1 to 2017.Q2.

Exchange-traded funds (ETFs) We use data on ETF flows, assets under management, and the share invested in US equities from Morningstar. The sample is quarterly from 2002.Q1 to 2017.Q2. We start in 2002 to ensure that we have a sufficient number of ETFs in our sample.

State and local pension funds We use data on returns, the share invested in equities and fixed income, total assets, and target holdings in equities and fixed income for state and local pension plans from the Center for Retirement Research at Boston College. The sample is from 2002.Q1 to 2018.Q4.

Hedge funds We use returns data on hedge funds from Hedge Fund Research, Inc. The data are monthly and start in January 1990. We include all indices that are available for the full sample. We also add the asset-weighted composite index, which is available since January 2008.

13F holdings We source the 13F filings from FactSet. The sample is from 1999.Q2 to 2017.Q2.

Macro-economic data We use quarterly data on real GDP growth from the St. Louis Federal Reserve Bank FRED database, series GDPC1.

Asset prices Data on returns with and without dividends are from the Center for Research in Security Prices at the University of Chicago, Booth School of Business. We use the monthly, value-weighted return with and without dividends to compute the monthly dividend payment.

Survey expectations We use survey expectations of returns from Gallup, as also used by Greenwood and Shleifer (2014), who use the fraction of investors who are bullish (optimistic or very optimistic) minus the fraction of investors who are bearish. We update their data, which starts in 1996.Q4, to 2018.Q4. The series has a couple of missing observations.

C.2 Sector-level data: Flow of Funds

We summarize the adjustments we make to the FoF data and the precise mapping to our model.

C.2.1 Data items

We use Corporate Equities (Table 223) for equities and for fixed income we take the sum of Treasury Securities (Table 210) and Corporate and Foreign Bonds (Table 213). We use unadjusted flows (FU) and, for the levels, we use the unadjusted market values when available (LM) and otherwise the estimated level (FL). Net issuances are initially equal to the total aggregate flow. The Flow of Funds revises historical data every quarter.

C.2.2 Notation and data definitions

Sectors are indexed by $i = 1, ..., I$, where $i = Foreign$ refers to the foreign sector. We observe holdings of equities, $\mathcal{E}_{it}$, Treasuries, Tr_{it}, and corporate bonds, C_{it}. We refer to the sum of Treasuries and corporate bonds as bonds, $B_{it} = Tr_{it} + C_{it}$. The flows corresponding to each asset class are denoted by ΔF^a_{it}, $a = \mathcal{E}, Tr, C, B$. Aggregate levels and flows omit the subscript i, implying, for instance, for equities $\sum_i \mathcal{E}_{it} = \mathcal{E}_t$ and for bonds $B_t = \sum_i B_{it}$. Lastly, the gross capital gain for equities is denoted by R^X_t and the return inclusive of dividend payments is denoted by R_t. We define the total assets of sector i as $W_{it} = \mathcal{E}_{it} + B_{it}$. Net issuances, $ni_t = \frac{NI_t}{\mathcal{E}_{t-1}}$, are based on equity markets.

In the FoF, equity flows are defined by $\Delta F^{\mathcal{E}}_{it} = \mathcal{E}_{it} - \mathcal{E}_{i,t-1} R^X_t$.[81] We assume in what follows that the securities are adjusted at the end of the quarter, $\Delta F^a_{it} = \Delta F^a_{it} = (\Delta Q^a_{it})\, P^a_t, a = \mathcal{E}, B$. The total per-period flow is $\Delta F_{it} = \Delta F^{\mathcal{E}}_{it} + \Delta F^B_{it}$ and in relative terms $\Delta f^a_{it} = \frac{\Delta F^a_{it}}{\mathcal{E}_{i,t-1}}, a = \mathcal{E}, B$. In the model, the cumulative flow is defined as

$$f_t = \sum_{s=0}^{t} \frac{\Delta F_s}{\bar{W}_s}.$$

The proportional per-period total flow is given by $\Delta f_{it} = \frac{\Delta F_{it}}{W_{i,t-1}}$. The equity shares are $S_{jt} = \frac{\mathcal{E}_{jt}}{\mathcal{E}_t}$. The relative change in equity demand, adjusted for price effects, is given by $\Delta q^{\mathcal{E}}_{jt} = f^{\mathcal{E}}_{jt} \left(R^X_t\right)^{-1} = \frac{\Delta Q^{\mathcal{E}}_{jt}}{Q^{\mathcal{E}}_{j,t-1}}$. The aggregate per-period flow measure is defined as $\Delta f_{St} = \sum_j S_{j,t-1} \Delta f_{jt}$.

In the remainder of this subsection, we summarize the adjustments we make to the raw data to account for measurement challenges in the data. However, in every step we make sure that the market clearing conditions for levels and flows hold.

C.2.3 Adjustment for foreign holdings of equity and corporate bonds

The FoF reports total flows and holdings of corporate equities and corporate bonds, including foreign assets held by US investors. As we are interested in measuring the flow into the US equity market, we adjust the holdings and flows for foreign positions. Unfortunately, we do not know the holdings and flows of foreign assets by sector, but we do know the aggregate positions across investors. We discuss our measurement approach in the context of equities, but we apply the same procedure to corporate bonds.[82]

[81] When possible, the FoF also follows this definition in other classes and has moved to market values for fixed income securities as well. However, in some cases, investors report holdings at book value for fixed income and no direct data on purchases are available, in which case flows are impacted by valuation effects.

[82] As US Treasuries are only issued in the US, obviously no adjustment is required for US Treasuries.

Let $\mathcal{E}_{it}^j$ be the equity holdings of sector i in period t for $j = D, F, T$, that is, the investment in domestic (D) and in foreign (F) securities as well as their total (T). We define the set of all US institutions by US. We define $x_{US,t} \coloneqq \sum_{i \in US} x_{it}$ for $x = \mathcal{E}, F$, that is, for equity levels and flows.

We start from the following identities

$$x_{it}^D + x_{it}^F = x_{it}^T, \tag{82}$$

$$\mathcal{E}_{it}^j = \mathcal{E}_{i,t-1}^j R_t^{X,j} + \Delta F_{it}^j, \tag{83}$$

where $R_t^{X,j}$ is the capital gain as before. We observe x_t^D, x_t^F, x_{it}^T for $x = \mathcal{E}, F$. We assume that the capital gain that different investors earn in the US is the same across investors (that is, $R_{it}^{X,D} = R_t^{X,D}$), and we make the same assumption for the capital gain on foreign investments (that is, $R_{it}^{X,F} = R_t^{X,F}$).

We assume for all US institutions, $i \in US$, that their equity holdings are split in the same way across foreign and domestic equities:

$$\mathcal{E}_{it}^D = \phi_t \mathcal{E}_{it}^T, \forall i \in US.$$

It then follows that

$$\phi_t = \frac{\mathcal{E}_{US,t}^D}{\mathcal{E}_{US,t}^T} = 1 - \frac{\mathcal{E}_{US,t}^F}{\mathcal{E}_{US,t}^T},$$

where $\mathcal{E}_{US,t}^F$ and $\mathcal{E}_{US,t}^T$ can directly be observed in the FoF. This measures ϕ_t.

For flows, we assume that

$$\Delta F_{it}^D = \mathcal{E}_{i,t-1}^D \eta_t^D + \phi_{t-1} \Delta F_{it}^T,$$

where η_t^D is a taste shock that we assume to be common across investors and impacts investors in proportion to their position in the previous period. Aggregating across all US institutions implies

$$\Delta F_{US,t}^D = \mathcal{E}_{US,t-1}^D \eta_t^D + \phi_{t-1} \Delta F_{US,t}^T,$$

implying

$$\eta_t^D = \frac{\Delta F_{US,t}^D - \phi_{t-1} \Delta F_{US,t}^T}{\mathcal{E}_{US,t-1}^D} = \frac{(1 - \phi_{t-1}) \Delta F_{US,t}^T - \Delta F_{US,t}^F}{\mathcal{E}_{US,t-1}^T - \mathcal{E}_{US,t-1}^F},$$

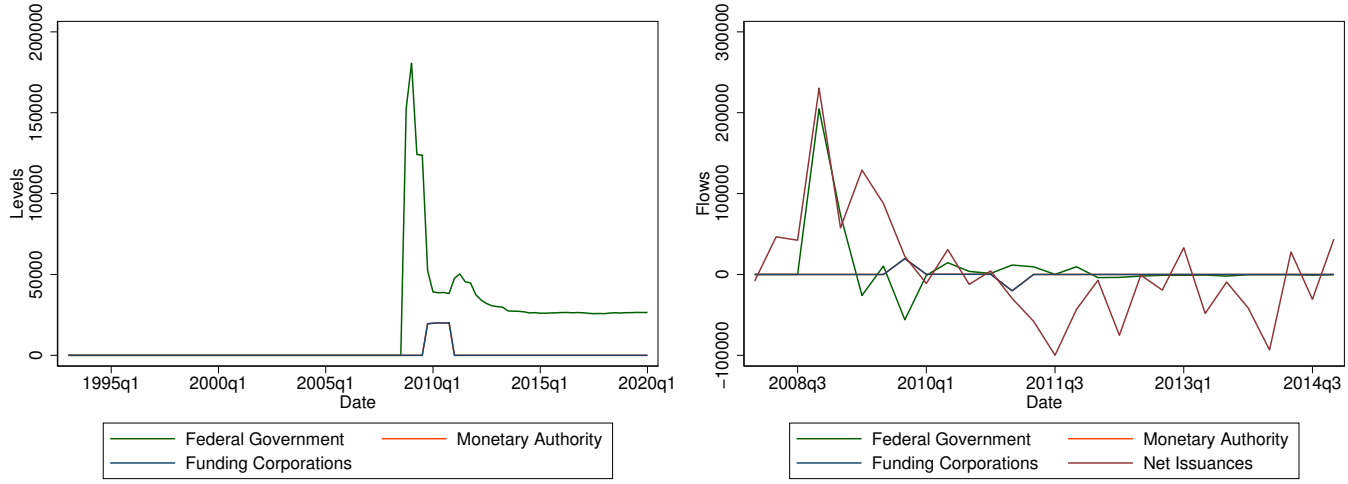
which can be computed directly from the FoF. With η_t and ϕ_t in hand, we can compute the estimate of domestic equity holdings and flows. We also adjust aggregate flows and levels to ensure that market clearing holds.

C.2.4 The impact of the 2008-2009 financial crisis

Three sectors have non-zero equity holdings only since the 2008-2009 financial crisis: the Federal Government (sector 31), the Monetary Authority (sector 50), and Funding Corporations (sector 71). These positions are all associated with the federal financial stabilization programs. We describe the adjustments we make to these series.

The holdings of the Federal Government are derived from "corporate equities issued by commercial banking under the federal financial stabilization programs," "corporate equities issued by funding corporations (AIG) under the federal financial stabilization programs," "corporate equities issued by bank-holding companies (GMAC) under the federal financial stabilization programs,"

Figure C.1: Equity levels and flows around the 2008-2009 financial crisis. The left panel shows the levels for the Federal government, the monetary authority, and funding corporations. The last two sectors have identical holdings and flows, and are therefore visually indistinguishable. The sample is from 1993.Q1 to 2018.Q4. The right panel reports the flows associated with the same sectors as well as net issuances from 2008.Q1 to 2014.Q4.



and "corporate equities issued by GSEs under the federal financial stabilization programs." From 2009.Q4 - 2011.Q1, Funding Corporations and the Monetary Authority record the exact same equity holdings. Their holdings are zero elsewhere. It is only a small position, and it comes from the way the AIG bailout was structured (per correspondence with economists at the FoF). The holdings are described as "Federal Reserve Bank of New York's Preferred Interests in AIA Aurora LLC and ALICO Holdings LLC." Both are life insurance subsidiaries of AIG.

The dynamics of the levels are plotted in the left panel of Figure C.1 from 1993.Q1 to 2018.Q4. The dynamics of net issuances alongside the flows associated with the three sector are plotted in the right panel of Figure C.1 from 2008.Q1 to 2014.Q4. The flows from funding corporations and the monetary authority are identical and cannot be distinguished visually. As can be seen from the graph, the stabilization programs created a spike in net issuances and these issuances are not absorbed by the typical investor sectors.

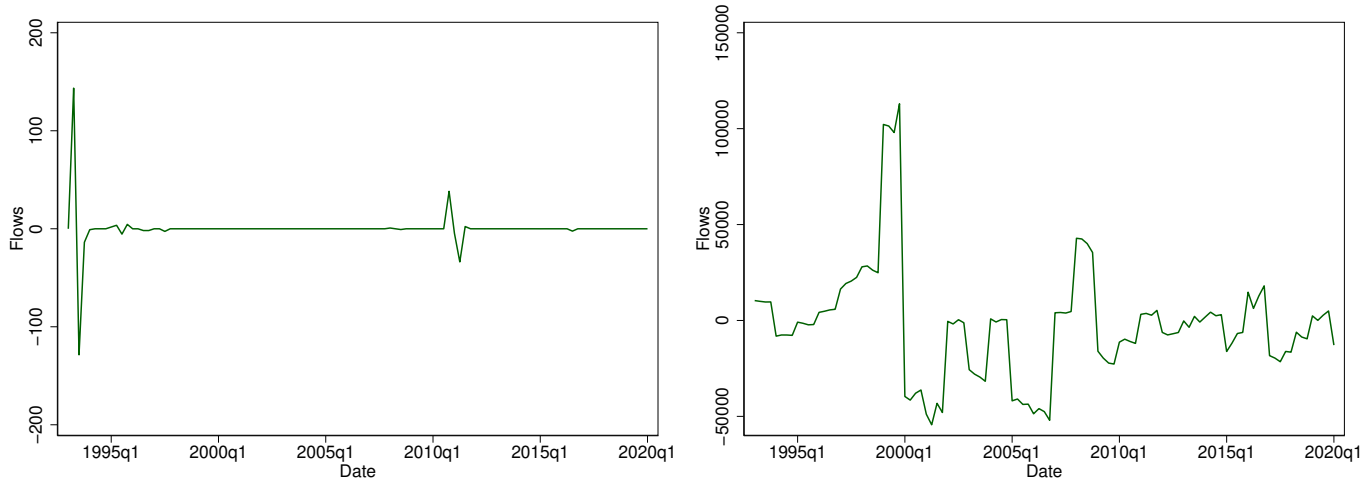
We aggregate the flows of these three sectors and subtract them from net issuances. We adjust the levels as well, and then remove these three sectors from our analysis.

C.2.5 Foreign banking offices in the US and non-financial corporate business holdings

We make adjustments for two additional sectors. First, the sector Foreign Banking Offices in the US (sector 75) has largely zero holdings since 1993, see the left panel of Figure C.2. Second, for the asset holdings of non-financial corporate businesses (sector 10), the quarterly flows are poorly measured, see the right panel of Figure C.2 showing the series from 1993.Q1 to 2018.Q4. The reason is that the FoF interpolates annual flows.[83] These flows and holdings reflect firms' holdings of other firms' equity, for instance for strategic or speculative reasons. Prior to the September 2018 publication, the FoF showed the equity liability of the non-financial corporate sector net of these

[83]For details of the current procedure, please see here.

Figure C.2: Flows of Foreign Banking Offices in the US and Non-financial Corporate Businesses



inter-corporate equity investments. The current release added the inter-corporate holdings as an asset and a liability. We undo this adjustment. For both sectors, we subtract the flows from net issuances and adjust the levels accordingly.

C.2.6 Examples of measurement issues

Even though the FoF data are the best data to use for both equity and fixed income holdings, certain measurement issues remain. We list them here and perhaps future research can refine some of our calculations. First, in the FoF, shares issued by ETFs, closed-end funds, and real estate investment trusts (REITS) are included in the corporate equities instrument category. This may impact the net issuance statistics, for instance. Most investor sectors do not separately report on ETFs, for instance, versus direct investments. As a result, we cannot adjust the holdings. Similarly, the total holdings include closely held equity. While the supply side is separated, we do not have disaggregated holdings, which implies we cannot adjust for this on the demand side.

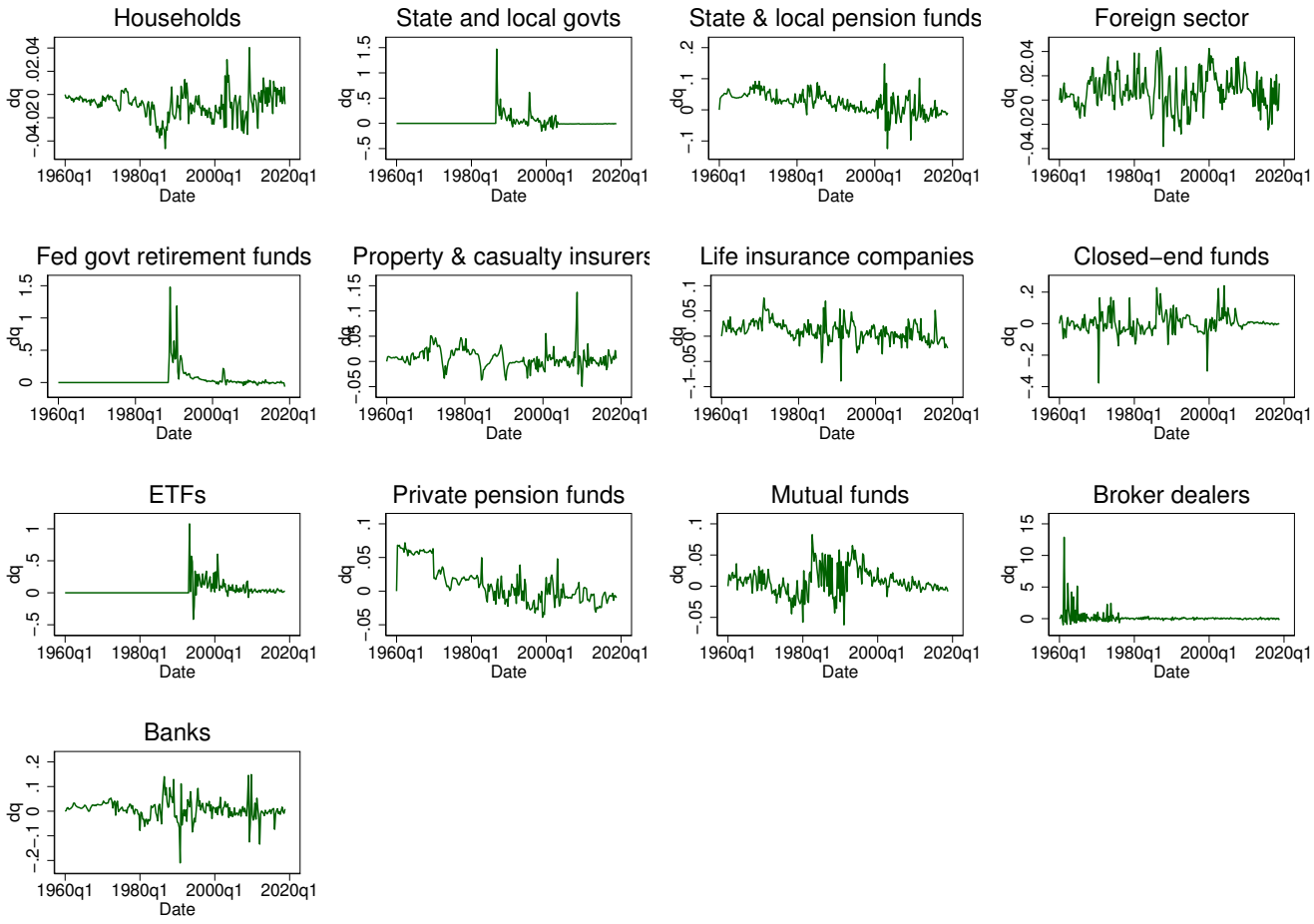
We start our sample in the early nineties. In part, this starting date is driven by the fact that institutional ownership has been rising, which allows us to provide a better disaggregation of the different sectors. However, the dynamics of equity flows, Δq, also looks more erratic in the earlier years. In Figure C.3, we plot the dynamics of equity flows across sectors to illustrate this issue. We therefore start in 1993. Lastly, ETFs become available in 1993. ETFs have been growing since then, in part to replace mutual funds, and we merge ETFs and mutual funds for parts of our analysis.

C.2.7 Sample construction of the data for the GIV estimation

Before implementing the GIV procedure, we make two adjustments to mitigate the impact of outliers. First, we merge the mutual fund and ETF sectors. ETFs were introduced in 1993, which is the start of our sample. The initial flows are very volatile, but their share of the overall market was small. This volatility gradually dissipates as the sector grows, in part at the expense of mutual funds. The volatility of the combined sector is much more stable over time.

Second, we winsorize the data by first removing the time-series median of each series, which is a robust way to remove differences in the levels of the series. We then winsorize each series

Figure C.3: Dynamics of equity flows across sectors. The figure shows the equity flows for the final 13 sectors in our sample from 1960.Q1 to 2018.Q4.



across time and sectors for the period from 1993.Q1 to 2006.Q4 to mitigate the influence of outliers. This avoids the need to winsorize during the financial crisis and the larger, as well as in the case of the more volatile flows happening during the earlier part of the sample. Winsorizing the data unconditionally does not impact our results much.

C.3 Investor-level data: 13F filings

We summarize the construction of the 13F data in this section.

C.3.1 Data construction

We source the 13F data from FactSet, version 5, which has also been used in Koijen et al. (2019). We start from the holdings data (table `own_inst_13f_detail_eq`) and we aggregate the holdings by roll-up entity (using table `own_ent_13f_combined_inst`). This combines filers that FactSet assigns as subsidiaries of the same investor. In addition, we aggregate the subsidiaries of BlackRock based on their names into a single entity. We identify an investor's type using table `own_ent_institutions`. We compute the market capitalization and holdings using the adjusted variables in FactSet (variables `adj_price`, `adj_shares_outstanding`, and `adj_mv`). In some rare cases, the total holdings exceeds the shares outstanding, which may be due to short-selling activity or filing errors. In these cases, we scale the holdings of all investors for a particular security to ensure that the market clearing condition holds. We then merge these data with the CRSP-Compustat merged data using CUSIPs. We select the securities with share code 10 or 11 and an exchange code equal to 1, 2 or 3. We use the capital gains data from CRSP. We use the CRSP-Compustat data to construct portfolio characteristics.

We also explore an alternative approach to aggregating filers. It is not uncommon for larger institutions to combine filings of subsidiaries into a single filing. When possible, FactSet dis-aggregates such filings, but preserves the mapping (table `own_ent_13f_subfiler_inst`). A potential concern is that a large institution moves assets between two of its subsidiaries, which is not an actual in- or outflow. We implement the following procedure to aggregate sub-filers as a robustness check. We start with the table `own_ent_13f_combined_inst`, which maps filers to parent entities. There are a couple of cases in which the parent entities themselves, and we roll them up one more level in those cases.[84] This leads to a new mapping of filers to roll-up entities.

Next, we use table `own_ent_13f_subfiler_inst`. We only consider those sub-filers (identified by `factset_subfiler_entity_id`) for which there is one filer (identified by `factset_filer_entity_id`). If an entity A files in some periods with entity B and in other periods with entity C, we keep it as a separate investor. As before, we roll-up the sub-filers to the ultimate filer. This results in a mapping from sub-filer to filer. We then combine with the mapping from filer to roll-up entities, and we obtain a new mapping from sub-filers to roll-up entities. These data are, by construction, more aggregated. Using this alternative approach, we find very similar results in terms of the multiplier estimates.

[84] That is, filer A rolls up into filer B, and filer B rolls up into filer C. There are also two cases in which the references are circular. That is, filer A rolls up to filer B and filer B rolls up to filer A. In those cases, we break the circle by replacing factset_rollup_entity_id to 05H9LL-E when the factset_rollup_entity_id equals 003RTX-E and replacing factset_rollup_entity_id to 06GK53-E when factset_rollup_entity_id equals 009F19-E.

C.3.2 Measuring changes in equity demand

We first discuss how we construct changes in equity demand, Δq_{it}. We denote by $H_{iat} = Q_{iat}P_{at}$ investor i's dollar holdings of security a at time t, where time t corresponds to the last day of the quarter. Total equity holdings are given by $\mathcal{E}_{it} = \sum_a H_{iat}$. We also define $\mathcal{E}_{it}^- = \sum_a H_{iat}^-$, where $H_{iat}^- = \frac{H_{iat}}{1+R_{at}^X}$, where R_{at}^X is the capital gain. In the absence of (reverse) splits, it holds $H_{iat}^- = Q_{iat}P_{a,t-1}$. We now define the flow as

$$\Delta q_{it} = 2\frac{\mathcal{E}_{it}^- - \mathcal{E}_{i,t-1}}{\mathcal{E}_{it}^- + \mathcal{E}_{i,t-1}},$$

which implies $\Delta q_{it} \in [-2, 2]$. This measure of flows is less sensitive to outliers than the alternative measure that uses only $\mathcal{E}_{i,t-1}$ in the denominator, see also Davis and Haltiwanger (1992). For consistency, we also use $\mathcal{E}_{it}^* = \frac{1}{2}\left(\mathcal{E}_{it}^- + \mathcal{E}_{i,t-1}\right)$ as our measure of size in what follows. Hence, $S_{i,t-1} = \frac{\mathcal{E}_{it}^*}{\sum_j \mathcal{E}_{jt}^*}$.

C.3.3 Constructing characteristics

Next, we discuss how we construct the characteristics that we use in Section B.2. We define the following characteristics

1. Log investor size, $\ln \mathcal{E}_{i,t-1}^*$.

2. Active share, which is defined as
$$\frac{1}{2}\sum_a \left|\theta_{iat} - \theta_{iat}^m\right|,$$
where θ_{iat} is the portfolio share and θ_{iat}^m the market-weighted portfolio of securities held by investor i at time t. We compute the eight-quarter median to mitigate the impact of outliers.

3. Three portfolio characteristics, namely size (as measured by market equity), value (as measured by the log market-to-book ratio), and market beta. In all cases, we compute the portfolio characteristic by computing $\sum_a \theta_{iat}c_{iat}$ for a given characteristic c_{iat}. For the size characteristic, we take the log value of the portfolio characteristic as the size distribution of firms is skewed. We compute the eight-quarter median to mitigate the impact of outliers.

4. Portfolio turnover, which is defined as
$$\sum_a \frac{\left|H_{iat}^- - H_{ia,t-1}\right|}{\mathcal{E}_{it}^*}.$$
As turnover is skewed, we use the log of 0.01 plus turnover as the characteristic. We the compute the eight-quarter median to mitigate the impact of outliers.

These characteristics define x_{it}, and we use their lagged values, $x_{i,t-1}$, to extract the factors. In addition to these characteristics, we also explored using investor type fixed effects, but they add little in terms of explanatory power beyond the characteristics and we therefore omit them.

C.3.4 Sample selection

We include an investor-quarter observation when we observe the flow and the lagged characteristics, $x_{i,t-1}$. Second, we remove investor-quarter observations where the flow in period t exceeds 100% and is below -100% in the next quarter (and vice versa). Such cases, although rare, may be data errors.[85] We drop investor-quarter observations in which investors hold fewer than 25 stocks or invest less than $50 million in equities for the securities in Section C.3.1. After these screens, we only include investors that satisfy these screens for at least 25 quarters. We impose the latter screen to ensure that we can accurately estimate the factor loadings and non-parametric loadings. In the final sample, we omit the household sector, which is measured as a residual.[86]We winsorize investor characteristics, each period, at the 2.5%- and 97.5%-percentiles, with the exception of size. Also, as with the FoF data, we winsorize Δq_{jt} (after removing the fund-specific median) at the 2.5%- and 97.5%-percentiles each period.

C.4 Mutual fund and ETF data

We follow the same procedure for US mutual funds and ETFs. We collect data on flows, assets, the equity share, and the US equity share. We omit fund-quarters in which the equity share, or the US equity share exceeds 300% or is lower than -300% as these may be data errors. We select the funds in Morningstar's US category group "US Equity" or "Allocation." We keep funds with a positive equity share and US equity share.

C.5 State and local pension fund data

State and local pension funds report once a year, but in different quarters. We use a fund's actual allocation to equity and fixed income and scale it so that the sum of the shares equals 100%. We also use the target allocations. In addition, each fund reports the return on the equity and fixed income portfolio, and we use the fund-specific returns.

D Additional empirical results

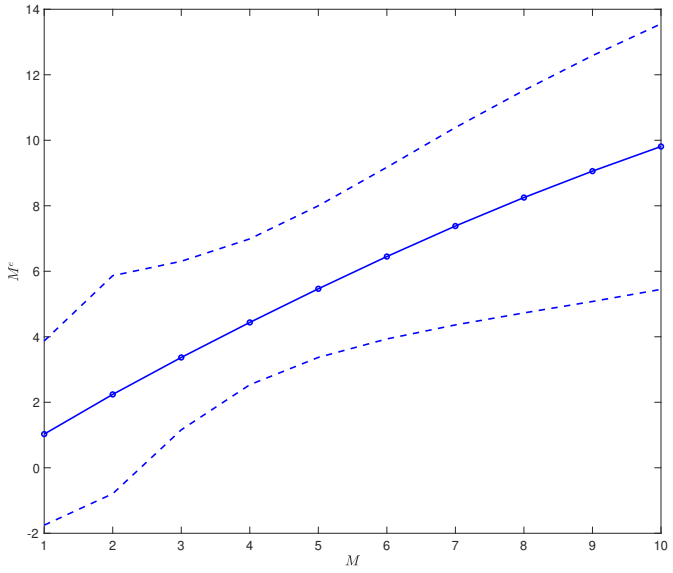
D.1 Performance of the GIV Estimator: Simulations

We illustrate the performance of the GIV estimator in an environment that closely mimics our empirical setting. We refer to Gabaix and Koijen (2020) for a more extended analysis of the GIV estimator. In particular, we proceed as follows to construct a realistic set of simulations. We start from our original sample and pick a value of ζ that we vary from $\zeta = 1$ (a multiplier of 1, a typical estimate for the micro elasticity) to $\zeta = 0.1$ (a multiplier of 10). For each value of ζ, we construct $f_t^\nu = \Delta q_t + \zeta \Delta p_t$. We then assume that the data follow a factor model and estimate $f_t^\nu = \lambda \eta_t + u_t$

[85]In addition, there is one small subsidiary of BlackRock that reports a large outlier for one quarter, and we remove this subsidiary with FactSet entity id equal to 0FSVG4-E. The same is true for a subsidiary of Putnam with FactSet entity id equal to 000V4B-E. Lastly, Janus Capital Group Inc. and Henderson Group plc merge in 2017.Q2 which leads to a large outlier that has not been updated in our version of FactSet and we drop the investor in 2017.Q2 (FactSet entity id equal to 000N8F-E).

[86]However, this also means that all measurement errors end up in the household sector. In addition, short-selling activity also distorts the measurement of the household sector as the 13F data only contains the long positions.

Figure D.4: Simulation results. The horizontal axis shows the multiplier in the data generating process and the vertical axis the average estimated multiplier, alongside the 2.5% and 97.5% percentiles, across 50,000 replications. We use the same sample size as in the empirical application in the next section. The text provides further details.



using principal components analysis. This provides us with an estimate of λ and V_u. These are the estimates one would obtain given the data we observed historically and if the true elasticity were equal to the assumed value. It tells us the volatility of aggregate shocks (μ_λ), the average volatility of idiosyncratic shocks ($\mu_{\ln\sigma}$), and the dispersion in these parameters across investors (σ_λ and $\sigma_{\ln\sigma}$).

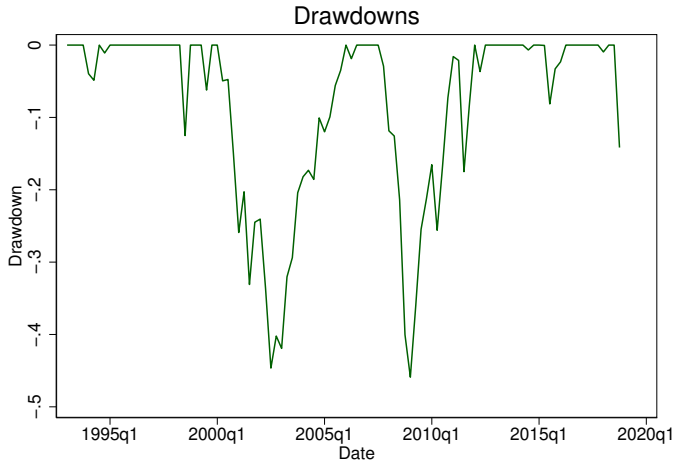
To simulate the data, we assume that $\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$, $\ln\sigma \sim (\mu_{\ln\sigma}, \sigma_{\ln\sigma}^2)$, and that the shocks are normally distributed. In doing so, we ensure that the volatility of prices is the same as in the data. Throughout, we use the same size distribution across sectors, which on average equals the one that we observe empirically. We then follow the standard procedure to estimate the multiplier.

We consider 50,000 replications for each value of $\zeta = 0.1, 0.2, \ldots, 1$ and report the average estimate alongside the 2.5% and 97.5% percentiles across all replications in Figure D.4. We report the multiplier M corresponding to the true data-generating process on the horizontal axis and the distribution of the estimated multipliers, M^e, on the vertical axis. The key takeaway is that our estimates uncover the true multipliers accurately with the dimensions of N and T that we observe empirically.

D.2 Drawdown dynamics

In Figure D.5, we plot the drawdowns, defined as the decline in the cumulative stock market index relative to its maximum so far, of the CRSP value-weighted index. We use these drawdowns to date recessions that we study in Section 2.

Figure D.5: The figure illustrates the drawdowns of the US stock market from 1993.Q1 to 2018.Q4. Drawdowns are defined as the ratio of the cumulative return index relative to its running maximum minus one.



D.3 Flows across investor classes are small

The flows across sectors are not only small during downturns, but also on average. To assess the magnitude of equity risk reallocation across sectors, we compute $y_t^{Gross} = \frac{\sum|\Delta F_{jt}^{\mathcal{E}}|+|\Delta F_t^{Firm}|}{2\mathcal{E}_{t-1}}$, where ΔF_t^{Firm} denotes net issuances of equity by firms. We divide the measure by two as for every buyer of \$1 of equity, there is a seller of the same amount. As some of the flows are associated with net repurchases, we separately measure the equity risk "creation" and "redemption" as a result of such corporate actions via $y_t^{AbsNet} = \frac{|\Delta F_t^{Firm}|}{\mathcal{E}_{t-1}}$, which we will refer to as absolute net flows.

The average absolute net flow equals 0.30% per quarter and the average gross flows average to 0.87% per quarter for the period from 1993.Q1 to 2018.Q4. The standard deviations are 0.26% and 0.37%, respectively. The difference between the series measures the risk reallocation in equity markets across institutional sectors, which averages to approximately 0.6% per quarter.[87] We plot the time series of both measures in Figure D.6 for the period from 1993.Q1 to 2018.Q4. The key takeaway is that the amount of equity risk that gets reallocated across sectors is small. These small flows contrast with the high levels of trading volume that are observed. However, much of this trading activity is at the single stock level, that is, exchanging stock A for stock B, instead of movements in or out of the stock market.

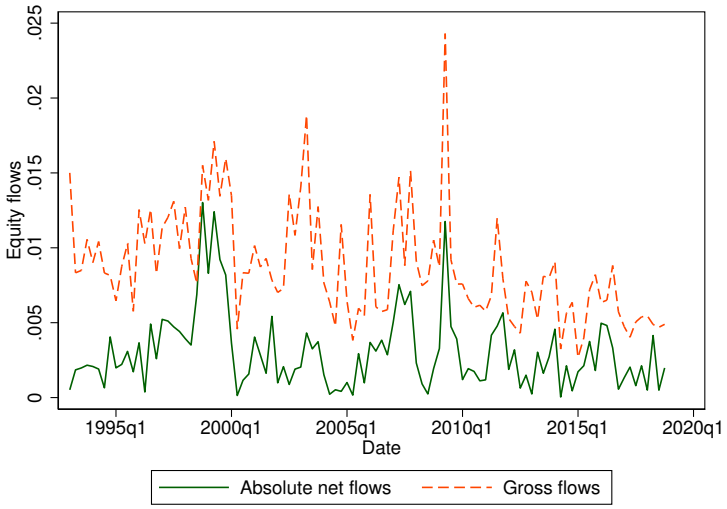
Small flows are not necessarily inconsistent with elastic markets. Many modern asset pricing models do not feature any trade. However, in the presence of volatile preference or belief shocks, this evidence implies that investors must experience the same shocks to preferences or beliefs, and have virtually the same exposure to these shocks, as otherwise we would see large flows across sectors.

In addition to quantities alone, Appendix F.1 also provides some additional first evidence on the link between flows and prices. Indeed, the demand by households (including mutual funds and

[87]This number is an upper bound to the extent that we care about the aggregate market elasticity as some of the flows between sectors are low-frequency time trends such as the shift from pension funds to mutual funds in the nineties or the shift from mutual funds to ETFs during the last twenty years.

ETFs) is positively correlated with price changes while the demand of the other sectors is strongly negatively correlated with price changes. This is consistent with the inelastic markets hypothesis in which shocks from the household sector, as defined by the FoF, lead to volatile prices as market are inelastic.

Figure D.6: The figure illustrates the reallocation of equity risk across various institutional sectors. The gross and net flow are defined in the main text. The sample is from 1993.Q1 to 2018.Q4.



D.4 Estimating κ using fund-level data

We explore the sensitivity of the equity share of various institutions to risk premia, which corresponds to κ in the model. To assess the response to risk premia, we consider the following simple regression as a starting point

$$\ln \theta_{it} = \alpha_i + \kappa \pi_t + \epsilon_{it}, \tag{84}$$

which we estimate using fund-level data for state and local pension plans, mutual funds, and ETFs. We discuss the data in Appendix C. As a measure of the equity premium, we consider a predictive regression

$$r^e_{t+4} = a + bDP_t + e_{t+4},$$

where r^e_{t+4} is the annual excess return and DP_t the dividend yield. We estimate the predictive regression using data from 1947.Q1 to 2018.Q4 and find a coefficient of $b = 3.8$.

We start with state and local pension funds and focus on the allocation between equity and fixed income from 2001 to 2018. We define θ_{it} as the fraction in equities relative to the fraction of total assets invested in equity and fixed income. For state and local pension plans, we also observe their target allocation, θ^T_{it}. We estimate (84) for both θ_{it} and θ^T_{it}. We have 2,957 fund-year observations for θ_{it} and 2,650 for θ^T_{it}.

We plot the equity-weighted average shares, alongside the equity premium, in Figure D.7. We adjust the mean and volatility of the equity premium to match the moments of θ_{St}.

Three observations stand out from the figure. First, the actual equity share drops during the global financial crisis. Second, the target equity share is more stable and, if anything, partially

Figure D.7: Equity share and target equity share of state and local pension plans. The figure shows the equity-weighted equity shares and target shares of state and local pension plans from 2001 to 2018. In addition, we plot the equity risk premium, which is constructed using a predictive regression of annual excess returns on the dividend yield from 1947.Q1 to 2018.Q4. We adjust the mean and volatility of the equity premium to match the moments of θ_{St}.



adjusts to the actual equity share.[88] Third, the shares are quite stable over time. The target share, which may be most relevant for the purposes of the calibration, hovers between 66% and 69%.

The estimates of $\kappa = -0.3$ (s.e. 0.8)[89] when using θ_{it} and $\kappa = 0.5$ (s.e. 0.2) when using θ_{it}^T, where the latter may be more representative. Of course, these are just OLS estimates for calibration purposes and primarily illustrate the lack of responsiveness of portfolio shares to equity premium variation.

We repeat this exercise for mutual funds. We restrict attention to US equity funds and allocation funds. For mutual funds, the sample runs from 1993.Q1 to 2017.Q2. We find again an insignificant, and close to zero, estimate for $\kappa = -0.1$ (s.e. 0.1), consistent with the stable equity shares in Figure 1. For ETFs, the equity-weighted equity share is essentially always equal to one (see Figure 1) and hence $\kappa = 0$. As for mutual funds, the point estimate is not significantly different from zero (-0.3 with a s.e. of 0.15).

Lastly, we briefly analyze a group of investors that may be prime candidates to provide elasticity, namely hedge funds. For hedge funds, we unfortunately do not have data on their equity shares. However, there are two reasons to believe that hedge funds provide limited elasticity to the aggregate stock market. First, their overall equity holdings decline during bad times (see Ben-David et al.

[88] We also estimate a simple adjustment model

$$\Delta\theta_{it} = a_i + \phi\left(\theta_{i,t-1} - \theta_{i,t-1}^T\right) + e_{it},$$

and find $\phi = -0.24$ and this estimate changes to $\phi = -0.23$ with time fixed effects. So pension funds close about a quarter of the gap each year. But, interestingly, the target also adjusts to close the gap, which also lowers the elasticity.

[89] In all cases, we cluster standard errors by year.

Table D.7: Portfolio dynamics of hedge funds. We report estimates of the style regression (85) for various hedge fund indices. The first column lists the HFRI index, the next two columns report the start and end date, followed by the alpha and its standard error, the estimate of θ_i and its standard error, and the estimate of κ_i and its standard error.

Index	Start date	End date	alpha	s.e.	θ_i	s.e.	$\theta_i\kappa$	s.e.
Asset Weighted Composite Index	2008m1	2018m12	0.08%	0.08%	0.27	0.02	-3.46	0.98
RV: Fixed Income-Convertible Arbitrage Index	1990m1	2018m12	0.26%	0.08%	0.21	0.02	2.55	0.70
ED: Distressed/Restructuring Index	1990m1	2018m12	0.45%	0.08%	0.25	0.02	0.59	0.70
Event-Driven (Total) Index	1990m1	2018m12	0.39%	0.06%	0.33	0.02	-0.04	0.56
Equity Hedge (Total) Index	1990m1	2018m12	0.40%	0.08%	0.48	0.02	-1.65	0.67
Emerging Markets (Total) Index	1990m1	2018m12	0.31%	0.15%	0.61	0.04	-0.92	1.33
Emerging Markets: Asia ex-Japan Index	1990m1	2018m12	0.20%	0.16%	0.55	0.04	-0.32	1.40
EH: Equity Market Neutral Index	1990m1	2018m12	0.24%	0.04%	0.06	0.01	-0.20	0.37
EH: Quantitative Directional Index	1990m1	2018m12	0.26%	0.10%	0.69	0.02	-2.32	0.83
RV: Multi-Strategy Index	1990m1	2018m12	0.28%	0.05%	0.16	0.01	1.35	0.44
RV: Fixed Income-Corporate Index	1990m1	2018m12	0.20%	0.08%	0.24	0.02	3.42	0.66
Fund of Funds Composite Index	1990m1	2018m12	0.18%	0.06%	0.23	0.01	-2.74	0.56
FOF: Conservative Index	1990m1	2018m12	0.15%	0.05%	0.14	0.01	-0.90	0.39
FOF: Diversified Index	1990m1	2018m12	0.16%	0.07%	0.23	0.02	-2.73	0.59
FOF: Market Defensive Index	1990m1	2018m12	0.31%	0.08%	0.03	0.02	-3.54	0.72
FOF: Strategic Index	1990m1	2018m12	0.27%	0.09%	0.36	0.02	-4.16	0.80
Fund Weighted Composite Index	1990m1	2018m12	0.34%	0.06%	0.36	0.01	-1.35	0.50
Fund Weighted Composite Index - CHF	1990m1	2018m12	0.22%	0.06%	0.36	0.01	-1.19	0.53
Fund Weighted Composite Index - GBP	1990m1	2018m12	0.43%	0.06%	0.36	0.01	-1.34	0.53
Fund Weighted Composite Index - JPY	1990m1	2018m12	0.15%	0.06%	0.36	0.01	-1.26	0.51
ED: Merger Arbitrage Index	1990m1	2018m12	0.30%	0.05%	0.14	0.01	0.65	0.42
Macro (Total) Index	1990m1	2018m12	0.46%	0.10%	0.17	0.02	-0.23	0.89
Macro: Systematic Diversified Index	1990m1	2018m12	0.39%	0.10%	0.21	0.02	-3.72	0.90
Relative Value (Total) Index	1990m1	2018m12	0.40%	0.05%	0.16	0.01	0.50	0.46

(2012)). This may be driven by outflows or risk constraints. Second, we can study their market betas using returns. We consider the regression

$$r^e_{it} = \alpha_i + (\theta_i + \theta_i\kappa_i\hat{\pi}_{t-1})\, r^e_{mt} + e_{it}, \tag{85}$$

where $\hat{\pi}_t$ is the de-meaned equity premium. We estimate the model using monthly data and report the results for a wide range of hedge fund styles. The results are presented in Table D.7. We find that θ_i averages to 0.3 (ranging from 0 to 0.7 across styles) and κ_i averages to -1 (ranging from -4.2 to 3.4). The first index is the asset-weighted average, the estimate of κ_i is significantly negative at $\kappa_i = -3.5$.

In summary, for these broad group of institutions, we find little evidence of large values of κ, that is, of elastic demand. In traditional models, we would expect that portfolio shares vary wildly as risk and risk premia move around, but we do not find it empirically. There are two groups we have not studied directly due to a lack of micro data, namely households and foreign investors. For foreign investors, though, we may be able to extrapolate from the results presented so far. As the foreign sector is comprised of similar institutions with similar mandates and risk constraints, these insights may carry over to those investors as well.

D.5 GIV estimates using FoF data: Robustness

In this section, we illustrate the robustness of our estimate of the multiplier, M. Given the small number of sectors in the FoF data, we focus on the case with a single principal component (in addition to a factor with uniform loadings). In Table D.8, we first repeat the benchmark estimate as a point of reference in the first column. In the second column, we omit all principal components. In the third column, we do not merge mutual funds and ETFs. In the fourth column, we do merge mutual funds and ETFs again, but now winsorize over the full sample from 1993.Q1 to 2018.Q4 instead of from 1993.Q1 to 2006.Q4 (and hence omit the financial crisis). In column five, we omit the time trend in (77). In column six, we control for the lagged value of Δq_{jt} in (77), while allowing for heterogeneous slope coefficients across sectors. In the seventh column, we start the sample in 2000.Q1. In the last two columns, we change the winsorization of the regression weights to $\frac{1.25}{N_t}$ (Column 8) or $\frac{1.75}{N_t}$ (Column 9), see the first step of the GIV algorithm. The main takeaway is that the multiplier estimates are quite stable across the various specifications. The estimates of the multiplier vary between 5.1 and 8.0.

Table D.8: Robustness of the GIV estimates. The table reports estimates of the multiplier under different measurement assumptions. We first repeat the benchmark estimate as a point of reference in the first column from 1993.Q1 to 2018.Q4. In the second column, we omit all principal components. In the third column, we do not merge mutual funds and ETFs. In the fourth column, we do merge mutual funds and ETFs again, but now winsorize over the full sample from 1993.Q1 to 2018.Q4 instead of from 1993.Q1 to 2006.Q4. In column five, we omit the time trend. In column six, we control for the lagged value of Δq_{jt} in (77), allowing for heterogeneous persistence coefficients across sectors. In the seventh column, we start the sample in 2000.Q1. In the last two columns, we change the winsorization of the regression weights to $\frac{1.25}{N_t}$ (Column 8) or $\frac{1.75}{N_t}$ (Column 9). Standard errors that account for autocorrelation are reported in parentheses.

	Δp	Δp	Δp	Δp	Δp	Δp	Δp	Δp	Δp
Z	7.08	8.00	6.94	7.65	6.63	6.77	6.79	7.17	5.12
	(1.86)	(1.24)	(1.48)	(1.37)	(1.17)	(2.18)	(2.02)	(1.18)	(2.33)
GDP growth	5.99	5.99	6.06	6.02	6.14	6.01	6.20	5.99	5.97
	(0.69)	(0.59)	(0.66)	(0.75)	(0.69)	(0.73)	(0.74)	(0.62)	(0.80)
η_1	21.06		22.09	15.59	32.54	25.24	-30.65	14.44	26.23
	(13.58)		(11.11)	(11.58)	(6.40)	(13.64)	(15.65)	(10.71)	(15.45)
Constant	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	104	104	104	104	104	103	76	104	104

In Table D.9, we replicate Table 1, but now adding the lagged value of Z_t. Across various specifications, Z_t has a small positive autocorrelation of approximately 10-15%. As is clear by comparing both sets of estimates, including a lag does not change the estimates in a meaningful way, and the lagged value of Z_t is in all cases insignificant.

In Table D.10, we start from the benchmark results in the previous table and add additional principal components. Given that the cross-section is small (we only have 12 sectors once we merge

Table D.9: Robustness of the GIV estimates for the Flow of Funds: Persistence in Z. The table replicates Table 1, but now adding the lagged value of Z_t. The sample is from 1993.Q1 to 2018.Q4. Standard errors are reported in parentheses.

	Δp	Δp	Δq_E	Δq_E	Δq_C	Δq_C
Z	7.41	5.62				
	(1.96)	(1.21)				
Δp			-0.12	-0.16	-0.01	-0.01
			(0.03)	(0.04)	(0.01)	(0.02)
Z (lag)	-1.59	-1.75	-0.30	-0.40	-0.02	-0.02
	(1.84)	(1.46)	(0.26)	(0.29)	(0.09)	(0.08)
GDP growth	6.41	6.42	0.59	0.86	0.22	0.24
	(0.93)	(0.90)	(0.25)	(0.24)	(0.13)	(0.12)
η_1	21.64	24.36	3.88	5.24	-0.72	-0.64
	(13.74)	(13.10)	(1.58)	(2.42)	(0.62)	(0.66)
η_2		30.11		5.13		0.28
		(6.02)		(1.16)		(0.80)
Constant	-0.01	-0.02	0.00	0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	103	103	103	103	103	103

mutual funds and ETFs), the data are not well suited to go beyond one or two principal components, unfortunately. Nevertheless, for transparency, we show the results up to five principal components for completeness in Table D.10. By adding additional principal components, idiosyncratic shocks will end up as factors, which makes it more challenging for us to identify the multiplier precisely. If we go beyond two principal components, the multiplier declines somewhat from 5.3 with two principal components to a range from 3.5 to 4.2 with three to five principal components.

D.6 The volatility of idiosyncratic demand shocks

In Table D.11, we report the standard deviation of sector-specific demand shocks, u_{it}, as constructed and used in Section 4.4.

D.7 Screening capital flows

In Table D.12, we report the slope estimates, β_j, of the regression in equation (45). In the first column we list the sector, in the second column whether we consider a flow to be mismeasured, and in the third column the estimate of β_j for that particular sector.

Table D.10: Robustness of the GIV estimates for the Flow of Funds: Additional principal components. The table adds principal components starting from a single principal component. The sample is from 1993.Q1 to 2018.Q4. Standard errors are reported in parentheses.

	Δp	Δp	Δp	Δp	Δp
Z	7.08	5.28	3.89	4.23	3.46
	(1.86)	(1.10)	(0.92)	(0.60)	(0.54)
GDP growth	5.99	5.97	5.96	5.96	5.96
	(0.69)	(0.67)	(0.64)	(0.51)	(0.51)
η_1	21.06	23.72	25.76	25.26	26.39
	(13.58)	(12.79)	(7.26)	(7.66)	(8.36)
η_2		29.95	32.56	31.92	33.36
		(6.54)	(5.37)	(5.35)	(5.93)
η_3			-25.57	-25.06	-26.21
			(5.57)	(5.20)	(5.16)
η_4				16.34	15.93
				(8.34)	(6.68)
η_5					-18.10
					(6.20)
Constant	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Observations	104	104	104	104	104

Table D.11: Volatility of idiosyncratic demand shocks by sector of the Flow of Funds. The table reports the volatility of idiosyncratic demand shocks by sector. We follow the procedure outlined in Section 4.4 to estimate the demand shocks. The sample is from 1993.Q1 to 2018.Q4.

Sector	$\overline{S}_{it}\sigma(u_{it})$	$\sigma(u_{it})$
Households	0.43	1.05
Mutual funds	0.22	0.92
Foreign sector	0.19	1.44
State & local pension funds	0.13	1.69
Private pension funds	0.12	1.22
Broker dealers	0.04	7.24
Life insurers	0.03	1.45
Property & casualty insurers	0.02	1.66
State and local govts	0.02	3.18
Closed-end funds	0.01	2.99
Fed govt retirement funds	0.01	2.20
Banks	0.01	3.17

Table D.12: Assessing the mismeasurement of capital flows. The table reports the slope coefficient of the regression in equation (45) to assess whether capital flows are mismeasured. We consider a flow to be correctly measured when β_j is significantly different from one. The sample is from 1993.Q1 to 2018.Q4.

Sector	Flows included	β_j	T-statistics for $H_0 : \beta_j = 1$
Households	0	0.47	9.11
State and local govts	1	0.49	1.95
State & local pension funds	1	1.04	0.73
Foreign sector	0	0.35	6.90
Fed govt retirement funds	0	-0.03	12.08
Property & casualty insurers	0	0.55	3.30
Life insurance companies	0	0.61	2.61
Closed-end funds	1	1.08	0.85
ETFs	1	1.01	0.93
Private pension funds	1	1.10	1.49
Mutual funds	1	0.98	0.53
Broker dealers	0	0.07	22.78
Banks	0	-0.06	10.67

D.8 Survey Details

We conducted three surveys. The first survey by putting out a request via Twitter (using the #econtwitter tag) to complete an online survey. In addition, we asked participants of an online seminar at VirtualFinance.org to complete the same survey – this latter audience being naturally more representative of the population of academic researchers in finance. Both surveys were conducted before the paper was available online and before the seminar on May 8. We launched the Twitter survey on May 7. We asked four questions:

1. If a fund buys $1 billion worth of US equities (permanently; it sells bonds to finance that position), slowly over a quarter, how much does the aggregate market value of equities change?

2. In response to the fund buying $1 billion over the quarter, some other investors need to sell. Who are the likely investors (by type) to sell their positions? (Pick at most two investor types).

   - Potential answers:
     (a) Hedge funds.
     (b) Mutual funds or ETFs.
     (c) Long-term investors such as pension funds and insurance companies.
     (d) Broker dealers.
     (e) Households.
     (f) Foreign investors (of any type).
     (g) Firms issuing new equity
     (h) Other [open text box]. We received hardly any additional sectors and will omit it from the discussion.

3. Since December 2019, did the equity risk premium:[90]

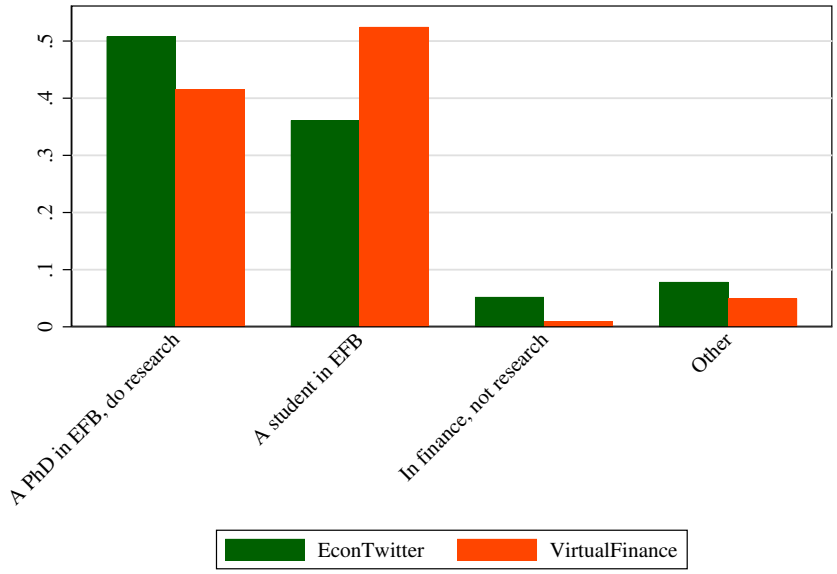
   - Potential answers:
     (a) Increase by more than 2.5%.
     (b) Increase between 0% and 2.5%.
     (c) Decrease between 0% and 2.5%.
     (d) Decrease by more than 2.5%.

4. Can you tell us a bit about yourself

   - Potential answers:
     (a) I am a student in economics / finance / business.
     (b) I have a PhD / doctorate in economics / finance / business, and do research.
     (c) None of the above.

[90] For the sake of brevity, we do not report the results for this question. It shows a significant amount of disagreement across respondents. In some models, such uncertainty about the exact value of the equity risk premium gives rise to inaction and therefor inelastic demand.

Figure D.8: Composition of survey respondents. The figure plots the distribution of respondents across two surveys; one conducted via Twitter (using the hashtag #EconTwitter) and one conducted at the beginning of a VirtualFinance.org seminar. The abbreviation EFB stands for Economics, Finance, and Business.



We also presented the paper a week later in the Virtual Macro Seminars (VMACS) on May 14. We repeated only the first and the last question, but attendees may have already seen the earlier presentation or have seen the slides. While the results are comparable, we consider it to be slightly polluted and focus on the earlier two surveys as a result. We remove responses that only signed the effect (e.g., "positive" or "negative" or ">0"). We received 192 responses via EconTwitter and 102 responses via VirtualFinance. In Figure D.8, we summarize the composition of responses. The abbreviation EFB stands for Economics, Finance, and Business. At least 85% of respondents are EFB students or have a PhD in EFB and do research.

In Table D.13, we summarize the responses about the multiplier, M. The main takeaway is that the profession views the aggregate stock market as highly elastic. Only 3% expects the multiplier to be larger than one and, in fact, fewer than 50% of the respondents expects a positive multiplier in each of the surveys. As a result the median multiplier estimate is zero in both surveys, and the mean is about 0.1. Note that this is even an order of magnitude smaller than the recent estimates of the micro elasticity of demand.

Given this feedback, it is interesting to explore the mechanism that may give rise to such high elasticities.[91] In Figure D.9, we provide the results to the third question, which points to hedge funds and broker dealers. We will explore these sectors in more detail in the paper. However, the bottom line is that broker dealers are fairly small as a sector and hedge funds do not appear to provide elasticity, and in particular not during economic downturns when the equity premium tends to rise sharply.

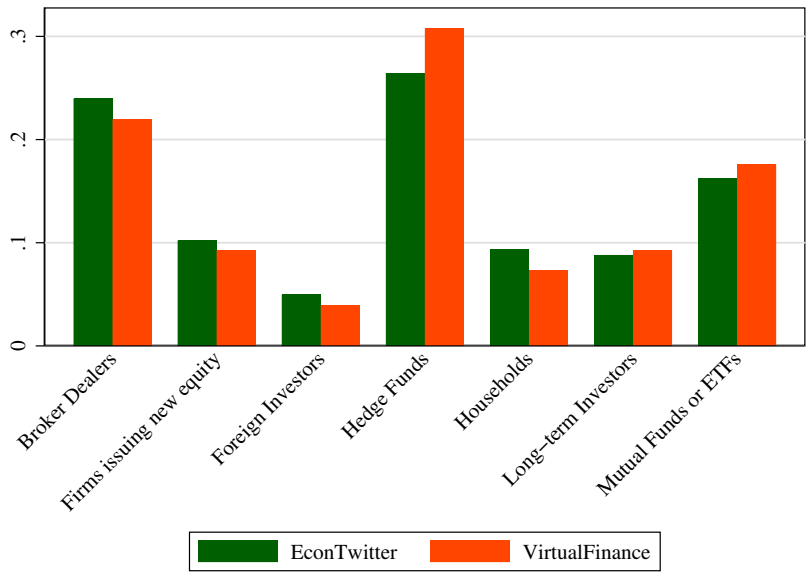
[91] This approach to "testing the mechanism" is similar in spirit to the ideas in Chinco et al. (2020) .

Table D.13: Survey responses regarding the multiplier. The table summarizes the distribution of survey responses about the multiplier, M. The data are from two surveys; one conducted via Twitter (using the hashtag #EconTwitter) and one conducted at the beginning of a VirtualFinance.org seminar. The first column reports the number of respondents. Columns 2 to 6 report the fraction of respondents who consider the multiplier to exceed one, be greater or equal to one, to exceed zero, equal to zero, or negative.

		Fraction of responses for which				
Survey	No. obs.	$M > 1$	$M \geq 1$	$M > 0$	$M = 0$	$M < 0$
VirtualFinance	102	2.9%	5.9%	47.1%	52.0%	1.0%
EconTwitter	192	3.1%	5.2%	29.5%	67.9%	2.6%

		Percentiles					
Survey	Mean	5	25	50	75	95	Max
VirtualFinance	0.13	0	0	0	0.01	1	5
EconTwitter	0.11	0	0	0	0	0.75	10

Figure D.9: Who provides elasticity to the market? The figure reports the fraction of respondents pointing to a particular sector as providing elasticity when an investors wants to sell $1bn worth of equities. The data are from two surveys; one conducted via Twitter (using the hashtag #EconTwitter) and one conducted at the beginning of a VirtualFinance.org seminar.



E Appendix: Theory Complements

E.1 Appendix: Omitted Proofs

Derivation of (14) We take a first order Taylor expansion of the expected return on equities, neglecting all variance and covariance terms as small:

$$\begin{aligned}
1+r+\bar{\pi}+\hat{\pi}_t &= 1+r+\pi_t \\
&= \frac{\mathbb{E}_t\left[P_{t+1}+D_{t+1}\right]}{P_t} \\
&= \frac{\mathbb{E}_t\left[\bar{P}_{t+1}\left(1+p_{t+1}\right)+\bar{D}_{t+1}\left(1+d_{t+1}\right)\right]}{\bar{P}_t\left(1+p_t\right)} \\
&= \mathbb{E}_t\left[\frac{\bar{P}_{t+1}}{\bar{P}_t}\left(1+p_{t+1}-p_t\right)+\frac{\bar{D}_{t+1}}{\bar{D}_t}\frac{\bar{D}_t}{\bar{P}_t}\left(1+d_{t+1}-p_t\right)\right] \\
&= \mathbb{E}_t\left[(1+g)\left(1+p_{t+1}-p_t\right)+(1+g)\,\delta\left(1+d_{t+1}-p_t\right)\right] \\
&= (1+g)\left(1+\delta\right)+(1+g)\,\mathbb{E}_t\left[\Delta p_{t+1}+\delta\left(d_{t+1}-p_t\right)\right]
\end{aligned}$$

The zero-th order term gives $1+r+\bar{\pi}=1+g+(1+g)\,\delta$, which is the Gordon growth formula, $r+\bar{\pi}-g=(1+g)\,\delta=\frac{\mathbb{E}_t[D_{t+1}]}{P_t}$. The next order term gives

$$\hat{\pi}_t=(1+g)\,\mathbb{E}_t\left[\Delta p_{t+1}+\delta\left(d_{t+1}-p_t\right)\right] \tag{86}$$

In the text, to reduce the notational clutter, we take 14, which is this expression with $g=0$. This can be interpreted as this expression in the limit of small time intervals, or when the trend growth rate g is 0, or by changing κ as $\kappa\,(1+g)$.

Proof of Proposition 6 *Case 1: Gaussian risk.* We first deal with the case of Gaussian risk, for simplicity in the continuous-time limit. The desired holding of risky wealth is $\theta_t=\frac{\pi_t}{\gamma\sigma^2}$. Initially, that holding was $\theta_t=1$: all wealth (human wealth and wealth capitalized in the stock market) is risky, with equal riskiness. This implies that $\gamma\sigma^2=\pi$ initially. But after the change in the equity premium, the desired change in equity share is: $d\theta=\frac{d\pi}{\gamma\sigma^2}$, i.e.

$$d\theta=\frac{d\pi}{\pi}. \tag{87}$$

The consumer can sell his wealth for P_t, so that his market wealth is $W_t=QP_t$, where Q is the total number of shares (of which $Q^{\mathcal{E}}$ are in equities, the rest in human wealth, i.e. promises to a stream of labor income). His dollar demand for risky assets is $W_t\theta_t$, so that in number of shares this is:

$$Q^D=\frac{W_t}{P_t}\theta_t=\frac{QP_t}{P_t}\theta_t=Q\theta_t=Q\left(1+\frac{\Delta\pi}{\pi}\right).$$

All the trading is in the equity market, so that this net demand for equities is:

$$\Delta Q=Q\frac{\Delta\pi}{\pi}.$$

This flow, expressed as a fraction of the equity market (which has a number of shares $Q^{\mathcal{E}} = \psi Q$), is also:

$$\frac{\Delta Q}{Q^{\mathcal{E}}} = \frac{\Delta \pi}{\psi \pi}. \tag{88}$$

If the value of equity changes by p, the equity premium changes by $\Delta \pi = -\delta p$ (see (14)), so we have

$$\frac{\Delta Q}{Q^{\mathcal{E}}} = -\frac{\delta}{\psi \pi} p = -\zeta^r p,$$

where the rational elasticity is:

$$\zeta^r = \frac{\delta}{\psi \pi}.$$

Finally, consumption is $C_t = Y_t$, while aggregate stock dividends are only $D_t Q^{\mathcal{E}} = \psi Y_t$. So,

$$\zeta^r = \frac{\delta}{\psi \pi} = \frac{\frac{D_t}{P_t}}{\frac{D_t Q^{\mathcal{E}}}{C_t} \pi} = \frac{C_t}{\left(P_t Q^{\mathcal{E}}\right) \pi} = \frac{C_t}{W_t^{\mathcal{E}} \pi},$$

which is the announced expression.

Case 2: Disaster risk. The reasoning is the same, except that expression (87) is different with disaster risk. To derive it, observe that the value function must take the form $V(W_t) = K \frac{W_t^{1-\gamma}}{1-\gamma}$ for some constant K. Hence, calling $\tilde{R}_{t+1}$ the rate of return on stocks, the consumer's problem is:

$$\max_{C,\theta} u(C) + \beta \mathbb{E} V\left(\left(W_t - C_t\right)\left(R_f + \theta\left(\tilde{R}_{t+1} - R_f\right)\right)\right)$$

It entails the following sub-problem for portfolio choice: $\max_\theta \mathbb{E}\left[\frac{\left(R_f + \theta\left(\tilde{R}_{t+1} - R_f\right)\right)^{1-\gamma}}{1-\gamma}\right]$. Calling $\tilde{r}_{t+1} = \frac{R_{t+1}}{R_f} - 1$ the normalized excess return on stocks, the problem is

$$\max_\theta \mathbb{E}\left[\frac{\left(1 + \theta \tilde{r}_{t+1}\right)^{1-\gamma}}{1-\gamma}\right],$$

so the FOC characterizing the equity share is:

$$\mathbb{E}\left[\left(1 + \theta \tilde{r}_{t+1}\right)^{-\gamma} \tilde{r}_{t+1}\right] = 0. \tag{89}$$

This expression holds for any i.i.d. excess return distribution $\tilde{r}_{t+1}$. In particular, it recovers the traditional expression $\theta = \frac{\pi}{\gamma \sigma^2}$ in the Gaussian case, $\tilde{r}_t = \pi \Delta t + \varepsilon_t$ (this is an exercise for the reader). Now take the disaster case,

$$\tilde{r}_t = \pi \Delta t - (1 - B) J_t$$

where π is the equity premium conditional on no disasters, where $J_t = 0$ if there is no disaster and 1 otherwise. Then (89) becomes

$$\left(1 - p^D \Delta t\right)\left(1 + \theta \pi \Delta t\right)^{-\gamma} \pi \Delta t + p^D \Delta t \left(1 + \theta\left(\pi \Delta t - (1 - B)\right)\right)^{-\gamma}\left(\pi \Delta t - (1 - B)\right) = 0,$$

i.e. taking the small $\Delta t \to 0$ limit,

$$\pi = p^D \left(1 - \theta(1 - B)\right)^{-\gamma} (1 - B). \tag{90}$$

Taking logs on both sides and differentiating this expression (for small changes in π and θ) around $\theta = 1$ gives:

$$\frac{d\pi}{\pi} = d\ln\pi = d\ln\left[p^D\left(1-\theta\left(1-B\right)\right)^{-\gamma}\left(1-B\right)\right] = \frac{\gamma\left(1-B\right)}{B}d\theta,$$

i.e.

$$d\theta = \frac{d\pi}{\pi}\frac{B}{\gamma\left(1-B\right)} \tag{91}$$

Finally, as $\pi = p^D B^{-\gamma}\left(1-B\right)$ by (90), the risk premium in a full sample (including an average number of disasters) is: $\bar{\pi} = \pi - p^D\left(1-B\right) = p^D\left(B^{-\gamma}-1\right)\left(1-B\right)$, so $\frac{\bar{\pi}}{\pi} = 1 - B^{\gamma}$. As $d\theta = \frac{d\pi}{\bar{\pi}}\frac{\bar{\pi}}{\pi}\frac{B}{\gamma(1-B)}$, this gives:

$$d\theta = \frac{d\pi}{\bar{\pi}}\frac{B\left(1-B^{\gamma}\right)}{\gamma\left(1-B\right)}, \tag{92}$$

which is the disaster counterpart to (87): how the desired equity share changes as the equity premium changes.

The rest of the derivation is exactly as in the lognormal Case 1, replacing (87) by (92).

In a behavioral model agents may have wrong beliefs, but they strongly act on their beliefs, with the same elasticity as in rational models (replacing the equity premium π by the perceived equity premium).

Proof of Proposition 8 Calling $\mathcal{D}_1$ the aggregate dividend, the dividend per share goes from $D_1 = \frac{\mathcal{D}_1}{Q_0}$ to $D_1' = \frac{\mathcal{D}_1}{Q_0'} = \frac{D_1}{1-b}$. So, the time-1 dividend per share increases by $d = b$.

Let us first consider a frictionless, elastic / rational model. The price per share increase by the same fraction as the time-1 dividend per share, i.e. $p = b$. Calling $v = \Delta\ln\left(PQ\right) = p + q^S$ the change in the market value of the firm, we have:

$$\text{Frictionless model: } q^S = -b, \qquad d = b, \qquad p = b, \qquad v = 0, \qquad r = 0.$$

The market value does not change: the lowering of the number of share outstanding by b is compensated by the increase in the price per share by a fraction, which is the same b.

Let us next consider an inelastic model. The buyback decreases the total dividend payout from D_0 to $D_0 - P_0Q_0b$, so "lowers the dividend from corporates to the fund by $f^{C\to M} = \frac{-P_0Q_0b}{W} = \frac{-\theta Wb}{W} = -\theta b$, so

$$f^{C\to M} = -\theta b.$$

Calling f^h the flow from households, the total flow is $f = f^h + f^{C\to M} = f^h - \theta b$. Hence, the "supply equal demand" condition in the share market translates into:

$$q^S = -b = q^D = -\zeta p + \kappa\delta d + f^h - \theta b,$$

so $\zeta p = \zeta b + f^h$ and

$$\text{Price change after buyback:} \quad p = b + \frac{f^h}{\zeta}. \tag{93}$$

We see that in the "frictionless limit" $\zeta \to \infty$, $p = b$, like in the rational model. But otherwise, it depends on f^h, the household flows.

We explore some modelling of f^h. Call μ^D (respectively μ^G) the fraction of the dividend (respectively, of the capital gain) that is "absorbed" by the households, that is, consumed, or reinvested

in bonds. Then, the primitive flow is $f^{h,0} = -\mu^D \theta \frac{\mathcal{D}'_0}{P_0 Q_0} - \mu^G \theta p$, so that the change in the flow is caused by the buyback is:

$$f^h = \mu^D \theta b - \mu^G \theta p.$$

For instance, the capital gain per share is p, the capital gain as a fraction of funds' assets is θp, and a fraction μ^G of that is "absorbed" (removed from the funds) by households. Hence, (93) gives:

$$p = \frac{\zeta + \mu^D \theta}{\zeta + \mu^G \theta} b. \tag{94}$$

This yields (66) and implies that $p > b$ if $\mu^D > \mu^G$. A share buyback increases the market value by $v = p + q^S = p - b > 0$.

Proof of Proposition 7 First, we derive the risk-free rate. The consumer's first order condition gives the Euler equation $1 = \beta R_{f,t} \mathbb{E}_t \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \right]$. As in equilibrium $C_t = Y_t$, the interest rate satisfies the Euler equation for bonds.

Now, we move to stocks. The average allocation in equities maximizes a risk-adjusted return, $\mathbb{E}_t \left[V^p \left(R_{t+1} \right) \right]$, with $V^p(R) = \frac{R^{1-\gamma}-1}{1-\gamma}$. Then, approximately, the allocation in equities is $\bar{\theta}^{\mathcal{E}} = \frac{\bar{\pi}}{\gamma \sigma_r^2}$. Given that in equilibrium all the wealth comes from equity, the equity premium is $\bar{\pi} = \gamma \sigma_r^2$.

Calling $W_t = W^* \left(\mathcal{D}_t \right) = \frac{1}{\theta} \frac{\mathcal{D}_t}{\delta}$ the value of the mixed fund when the behavioral friction is 0 (see footnote 62), (52) gives:

$$\Delta f_t = \frac{\Delta F_t}{W^*_{t-1}} = \frac{\frac{1-\theta}{\theta} \frac{Q^{\mathcal{E}} \Delta \mathcal{D}_t}{\delta} + \Delta \left(b_t \frac{Q^{\mathcal{E}} \mathcal{D}_t}{\delta} \right)}{\frac{1}{\theta \delta} \mathcal{D}_{t-1}}$$

i.e., using $\mathcal{D}_t = Q^{\mathcal{E}} D_t$,

$$\Delta f_t = (1-\theta) \frac{\Delta D_t}{D_{t-1}} + \theta \frac{\Delta \left(b_t D_t \right)}{D_{t-1}} \tag{95}$$

hence

$$f_t = (1-\theta) d_t + \tilde{f}_t, \tag{96}$$

where $d_t = \sum_{s=1}^{t} \frac{\Delta D_s}{D_{s-1}}$ is the cumulative increase in the dividend and, to the leading order (for b_t close to 0, and in the limit of small time intervals where $\frac{D}{D_{t-1}}$) $\tilde{f}_t = \theta b_t$.

The expressions for price and the equity premium were derived in (22). Note that what we call p_t here the value of $p_t - d_t$ in (22). The rest (e.g. expressions for ζ) is derived in the paper.

E.2 Different asset classes

We can easily extend the model to K asset classes, indexed by $A \in \{1, \ldots, K\}$, such as stocks, long-term government bonds, and long-term corporate bonds. This way, we can study cross-market contagion effects, and the impact of those on real investment. We sketch this for the two-period model of Section 3.1.

The mandate leads to the following demand for asset A (at least, for some small deviations from 0 in d and p):

$$P_A Q_A^D = \theta_A W \exp \left(\sum_{B=1}^{K} \kappa_{AB}^D \left(d_B - p_B \right) \right).$$

For instance, if $\kappa^D_{AB} = 0$ the mixed fund seeks to keep a constant share θ_A in asset A. When κ^D_{AB} is different from 0, a change in the risk premium in asset B leads to a change in the amount allocated to asset A.

Suppose that there are shocks changing the prices and expected dividends for a given set of assets by fractions indexed as p_B and d_B. Then, the value of the fund changes by $w = \frac{\Delta W}{W} = f + \sum_B \theta_B p_B$, so that the demand for a particular asset A class changes by a fraction

$$q^D_A = -\sum_B \zeta_{AB} p_B + f_A + \sum_B \kappa^D_{AB} d_B, \tag{97}$$

where the cross-elasticities of demand ζ_{AB} express how demand for asset A changes with a change in the price of asset B: $\zeta_{AB} = 1_{A=B} - \theta_B + \kappa^D_{AB}$. In vector form, this gives

$$q = -\zeta p + f + \kappa^D d, \tag{98}$$

where now q, p, d, f are vectors, and ζ and κ^D are matrices, with dimension K. This generalizes Proposition 2. So, the equilibrium after a change in flows and expected dividends (but still constant asset supply is):

$$p = \zeta^{-1} \left(f + \kappa^D d \right). \tag{99}$$

In this paper we shall not measure, for example, how much the price of long-term bonds affects the demand for stocks. But one can readily contemplate a host of interesting cross-market effects. For instance, when investors sell stocks and invest in long-term bonds, bond yields will go down, which encourages firms to invest. Hence, we see an impact from stocks to corporate bonds, to real investment, and to GDP.

E.3 The model in continuous time

E.3.1 Some of the main expressions, in continuous time

We use the notation $\mathbb{E}_t \left[\frac{dp_t}{dt} \right] = \frac{\mathbb{E}_t[dp_t]}{dt} := \lim_{h \downarrow 0} \mathbb{E}_t \left[\frac{p_{t+h} - p_t}{h} \right]$. So, if $dp_t = \mu_t dt + \sigma_t dZ_t$, then $\mathbb{E}_t \left[\frac{dp_t}{dt} \right] = \mu_t$. Here we record the main expressions in continuous time. The equity premium

$$\pi_t = \mathbb{E}_t \left[\frac{dP_t}{P_t} \right] / dt + \frac{D_t}{P_t} - r \tag{100}$$

has the Taylor expansion:

$$\hat{\pi}_t = \mathbb{E}_t \frac{dp_t}{dt} + \delta \left(d_t - p_t \right). \tag{101}$$

We have

$$q^D_t = -\zeta p_t + \kappa \delta d_t + \mathbb{E}_t \left[\frac{dp_t}{dt} \right] + \nu_t + f_t,$$

which in equilibrium (with $q^D_t = 0$) leads to the stock price equation

$$\mathbb{E}_t \left[\frac{dp_t}{dt} \right] - \rho p_t + \delta d_t + \frac{\nu_t + f_t}{\kappa} = 0. \tag{102}$$

Integrating forward, the stock price is:

$$p_t = \mathbb{E}_t \int_t^\infty e^{-\rho(\tau-t)} \left(\frac{\rho}{\zeta} (f_\tau + \nu_\tau) + \delta d_\tau \right) d\tau \tag{103}$$

$$= \frac{\delta}{\rho} d_t + \frac{1}{\zeta} (f_t + \nu_t) + \mathbb{E}_t \int_t^\infty e^{-\rho(\tau-t)} \left(\frac{\dot{f}_\tau + \dot{\nu}_\tau}{\zeta} + \frac{\delta}{\rho} \dot{d}_\tau \right) d\tau. \tag{104}$$

This allows for easier calculations than the discrete time model. For instance, suppose that flows and dividends follow autoregressive process, e.g. $df_t = -\phi_f f_t + \sigma_t dZ_t$ (for dZ_t a mean-zero increment process, e.g. a Brownian motion). Then we have $\mathbb{E}_t [f_\tau] = e^{-\phi_f(\tau-t)} f_t$ for $\tau \geq t$, so (103) gives:

$$p_t = \frac{\rho}{\rho + \phi_f} \frac{f_t}{\zeta} + \frac{\delta}{\rho + \phi_d} d_t = \frac{1}{\zeta + \kappa\phi_f} f_t + \frac{\delta\kappa}{\zeta + \kappa\phi_d} d_t. \tag{105}$$

which is the continuous time equivalent of (22).

Combining (105) and (18) leads to:

$$\hat{\pi}_t = b_f^\pi f_t + b_d^\pi d_t, \tag{106}$$

with $b_f^\pi < 0$ and $b_d^\pi > 0$. In the random walk case, $b_f^\pi = -\frac{\delta}{\zeta}$ and $b_d^\pi = \frac{\delta(1-\theta)}{\zeta}$, while in the general case, we have $b_f^\pi = -\frac{(\delta+\phi_f)\rho}{\rho+\phi_f}\frac{1}{\zeta}$ and $b_d^\pi = \frac{\delta(1-\theta)}{\zeta+\kappa\phi_d}$.

E.4 Micro versus macro elasticity: The cross-section of stocks

We generalize to a model with several stocks. This allows us to distinguish between the macro elasticity of demand for stocks, ζ and the micro-elasticity $\zeta^\perp$. The upshot is that the effects are the same, but with higher demand elasticity in the cross-section $\zeta^\perp > \zeta$ than in the aggregate. We recommend skipping this section at the first reading.

E.4.1 Stock-level demands

We call P_{at} the price of the stock, and p_{at} its deviation from the baseline (as we did for the aggregate market). We define $p_a^\perp = p_a - p$ as the asset-a specific price deviation. Likewise, all "perpendicular" terms are the deviation of stock a from the aggregate stock market. We define $\pi_{at}^\perp = \pi_t^a - \beta_a \pi_t$ as the deviation of the equity premium of asset a from the CAPM benchmark (this could be generalized of course), and $\hat{\pi}_{at}^\perp = \pi_{at}^\perp - \bar{\pi}_a^\perp$ as its deviation from the average.

We start from a model of stock-level demand for stock a (as in *a*sset), which comes from a "tracking error" type of mandate: the fraction in equities allocated to asset a is

$$\frac{P_{at} Q_{at}^D}{P_t Q_t^D} = \theta_a^{\mathcal{E}} e^{\kappa^\perp \hat{\pi}_{at}^\perp + \theta^\perp p_a^\perp}. \tag{107}$$

Indeed, $P_{at}Q_{at}^D$ is the dollar demand for asset a, and $P_t Q_t^D$ is the dollar demand for the aggregate stock market. On average, their ratio is $\theta_a^{\mathcal{E}}$. The term $\kappa^\perp$ is the micro-elasticity of demand with respect to the anomalous part of the equity premium $\hat{\pi}_{at}^\perp$. The term $\theta^\perp$ indicates a concern for

tracking error: if the fraction allocated to asset a is constant, then $\theta^{\perp} = 0$ (this is the baseline case). However, if the number of shares allocated to asset a is constant, then $\theta^{\perp} = 1$.

Calling $q_{at}^{d} = \frac{Q_{at}}{Q_a} - 1$ the deviation of the demand from the baseline, and $q_{at}^{D,\perp} = q_{at}^{D} - q_t^{D}$ how much asset a deviates from the baseline, we obtain the following counterpart to Proposition 4 (the proof is in Appendix E.1).

Proposition 10. *(Demand for individual stocks in the infinite-horizon model) The demand change (compared to the baseline) for an individual asset a is $q_{at}^{D} = q_t^{D} + q_{a,t}^{D,\perp}$, where q_t^{D} is the demand change for the aggregate stock market seen in Proposition 4, and $q_{at}^{D,\perp}$ is the asset-a specific demand change, given by*

$$q_{at}^{D,\perp} = -\zeta^{\perp} p_{at}^{\perp} + \kappa^{\perp} \delta d_{at}^{e,\perp} + \kappa^{\perp} \mathbb{E}_t \left[\Delta p_{a,t+1}^{\perp} \right] \tag{108}$$

where $\zeta^{\perp}$ is the micro-elasticity of demand for individual stocks:

$$\zeta^{\perp} = 1 - \theta^{\perp} + \kappa^{\perp} \delta. \tag{109}$$

This is exactly the same equation as the one for the aggregate stock market, but now in terms of stock-specific deviations. Hence, the economics of the aggregate stock market works for the individual stocks, but in "perpendicular space", i.e. replacing ζ, p_t, q_t^{D} by $\zeta^{\perp}$, $p_{at}^{\perp}$, $q_{at}^{D,\perp}$, and so on. We next spell this out and draw consequences.

Proof of Proposition 10 Equation (107) implies

$$q_{at}^{D,\perp} = -\left(1 - \theta^{\perp}\right) p_{at}^{\perp} + \kappa^{\perp} \hat{\pi}_{at}^{\perp} + \nu_{at}^{\perp}. \tag{110}$$

Likewise, the analogue of (14) is $\hat{\pi}_{at}^{\perp} = \mathbb{E}_t \left[\Delta p_{a,t+1}^{\perp} \right] + \delta \left(d_{at}^{e,\perp} - p_a^{\perp} \right)$, with $d_{at}^{e,\perp} := \mathbb{E}_t \left[d_{a,t+1}^{\perp} \right]$. Combining the two gives the announced expression. □

E.4.2 Micro-elasticity of demand versus macro-elasticity of demand

Suppose that there is a "stock specific flow", whereby someone buys $\Delta F_a^{\perp}$ worth of stock a, while selling $\Delta F_a^{\perp}$ of the aggregate stock market, so that the total change in the demand for aggregate stocks is 0. The asset-a specific fractional inflow is $f_a^{\perp} = \frac{\Delta F_a^{\perp}}{P_a Q_a}$, where $P_a Q_a$ is the (pre-flow) market value of stock a. As net demand is 0, we must have $q_{at}^{D,\perp} + f_a^{\perp} = 0$. So, the impact of a flow is:

$$p_a^{\perp} = \frac{f_a^{\perp}}{\zeta^{\perp}}, \tag{111}$$

where $\zeta^{\perp}$ is the price micro-elasticity of demand (109). We see that the price impact is $\frac{1}{\zeta^{\perp}}$, not $\frac{1}{\zeta}$.

Calibration Most papers have estimated the micro-elasticity of demand, $\zeta^{\perp}$ (Shleifer (1986), Wurgler and Zhuravskaya (2002), Duffie (2010), Chang et al. (2014), Koijen and Yogo (2019)), while the present paper is about the macro-elasticity of demand, ζ. Indeed, the literature finds $\zeta^{\perp} \simeq 1$, with estimates in the 0.5 to 10 range. It makes sense that the macro-elasticity should be much smaller than the micro-elasticity, $\zeta \ll \zeta^{\perp}$. One way to rationalize this is to set $\theta^{\perp} \simeq 0.2$ for the inertia or concern for tracking error term, $\delta = 4\%$, and $\kappa^{\perp} = 5$.[92]

[92] This ratio of price impact of roughly 1 to 5 is also consistent with Benzaquen et al. (2017).

Micro versus macro price impact In the following illustrations, we take a micro elasticity $\zeta^{\perp} = 1$ and a macro elasticity $\zeta = 0.2$.

Consider what happens if an investor decides to buy \$1 worth of Apple shares, while selling \$1 worth of Google shares. Then, the market value of Apple goes up by \$1 (that is, $\$1 \times \frac{1}{\zeta^{\perp}}$), and that of Google falls by the same \$1. But the aggregate value of equities does not change, as the net demand for aggregate equities has not changed.

Next, suppose that an investor buys \$1 of a very small stock (selling \$1 worth of bonds), call it Peanut. Then, the market value of that Peanut stock goes up by \$1, and the market value of the aggregate stock market goes up by \$5 – so the aggregate market value of the other stocks increases by \$4.

If the consumer buys \$1 of Apple, or any non-infinitesimal stock, the aggregate value of equities still increases by \$5, but the market value of Apple goes up by slightly more than \$1 (indeed, if Apple were the whole market, its would increase by \$5). To see all this analytically, consider a flow $f_a = \frac{\Delta F_a}{P_a Q_a}$ into just one asset a, which accounts for a fraction ω_a of the total equity capitalization. We do that in the two-period model, so we drop t (this is equivalent to doing that for the infinite-horizon model, but assuming permanent inflows). The corresponding aggregate flow is $f = \omega_a f_a$, so that the impact on the aggregate market is $p = \frac{f}{\zeta}$, or

$$p = \frac{\omega_a f_a}{\zeta}.$$

The stock-specific flow to asset a is $f_a^{\perp} = f_a - f = (1 - \omega_a) f_a$. Hence, the stock-specific impact is: $p_a^{\perp} = \frac{f_a^{\perp}}{\zeta^{\perp}} = \frac{1-\omega_a}{\zeta^{\perp}} f_a$. Hence, the total impact is $p_a = p + p_a^{\perp}$, or

$$p_a = \frac{f_a}{\zeta^{\perp}} + \left(\frac{1}{\zeta} - \frac{1}{\zeta^{\perp}}\right) \omega_a f_a. \tag{112}$$

For the other stocks $b \neq a$, we have $f_b^{\perp} = -f = -\omega_a f_a$, so the impact is:

$$p_b = \left(\frac{1}{\zeta} - \frac{1}{\zeta^{\perp}}\right) \omega_a f_a, \qquad \text{for a stock } b \neq a. \tag{113}$$

As $\zeta < \zeta^{\perp}$, the cross-impact is positive.

For instance, suppose that Apple's capitalization is $\omega_a = 5\%$ of the stock market. Then, if someone buys \$1 of Apple (selling bonds), the market value of Apple increases by \$1.2,[93] and the value of the aggregate equities still increases by \$5 – so, the aggregate value of all the other stocks increases by \$3.8. This is a moderate deviation from the above "small stock" Peanut benchmark.

Infinite-horizon model for the cross-section The infinite horizon model is exactly as above, but in "perpendicular" (asset-specific) space. We define $\rho^{\perp} = \frac{\zeta^{\perp}}{\kappa^{\perp}} = \frac{1-\theta^{\perp}}{\kappa^{\perp}} + \delta$ and $M^{D,\perp} = \frac{\kappa^{\perp}\delta}{1-\theta^{\perp}+\kappa^{\perp}\delta}$. The stock-specific deviation is given by (16) in asset-specific space:

$$p_{a,t}^{\perp} = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho^{\perp}}{(1+\rho^{\perp})^{\tau-t+1}} \left(\frac{f_{a\tau}^{\perp}}{\zeta^{\perp}} + M^{D,\perp} d_{a\tau}^{e\perp}\right). \tag{114}$$

[93] Indeed, $\frac{1}{\zeta^{\perp}} + \left(\frac{1}{\zeta} - \frac{1}{\zeta^{\perp}}\right) \times \omega_a = 1 + (5-1) \times 5\% = 1.2$

Conclusion: Aggregate versus cross-section We conclude that the aggregate model transposes well to the cross-section, and indeed is useful to think about the impact of flows in the cross-section and in the aggregate in a unified manner. While most prior work has been on the estimation of the cross-sectional elasticity $\zeta^{\perp}$, the main object of interest in this study is the aggregate elasticity ζ.

E.5 The model with time-varying market inelasticity

Here we study the model with a time-varying market elasticity.

Suppose we have a time-varying ζ_t and κ_t but (for simplicity), a constant $\rho = \frac{\zeta_t}{\kappa_t}$. For simplicity, we assume $\mathbb{E}_t d^e_\tau = 0$. Then, we have the following variant of Proposition 5 (the derivation is similar):

$$p_t = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho}{(1+\rho)^{\tau-t+1}} \frac{f_\tau + \nu_\tau}{\zeta_\tau} \tag{115}$$

To be concrete, we study the case

$$\frac{1}{\zeta_t} = \frac{1}{\zeta}(1 + \mathcal{M}_t), \qquad \mathbb{E}_t \mathcal{M}_{t+1} = (1 - \phi_\zeta)\mathcal{M}_t,$$

so that $\mathcal{M}_t$ is a temporary increase in market inelasticity, mean-reverting at a speed ϕ_ζ. We consider the impact of a permanent inflow, $f_\tau = f_0$ for $\tau \geq 0$. Then the price follows:

$$p_t = \frac{f_0}{\zeta}\left(1 + \frac{\rho}{\rho + \phi_\zeta}\mathcal{M}_t\right). \tag{116}$$

So, if the flow f_0 happens during a time of high market inelasticity $\frac{1}{\zeta_t}$ (i.e. high $\mathcal{M}_t$), then the price impact is higher, which makes sense. It is the average future value of the inelasticity ($\frac{\rho}{\rho+\phi_\zeta}\mathcal{M}_0$) that matters, rather than the current inelasticity ($\mathcal{M}_0$). In the scenario above, the price impact of f_0 mean-reverts at a speed ϕ_ζ.

More generally (if $\phi_\zeta > \phi_f$), this implies implies that returns that happened during a high-volatility period mean-revert faster.

A tentative calibration. With $\rho = 0.13/year$ and $\phi_\zeta = 0.15/year$, we have $\frac{\rho}{\rho+\phi_\zeta} \simeq 0.4$, so have then to get a price impact higher by a factor 0.5, we need $\mathcal{M}_t = \frac{0.5}{0.4} = 1.2$, i.e. a halving of ζ_t. This effect might be detectable, though not easily.

E.6 When flows react to the risk premia

Here we derive Proposition 9, and more generally explore the consequences of flows of the type:

$$\Delta f_t = \chi \hat{\pi}_t + \varepsilon_t. \tag{117}$$

We first proceed in continuous time, which is cleanest.

E.6.1 Continuous time

For simplicity, we assume away dividend surprises. They would be easy to add back. The flows (117) are

$$df_t = \chi \hat{\pi}_t dt + \sigma dz_t. \tag{118}$$

We use the operator D,

$$Dx_t \coloneqq \frac{\mathbb{E}_t [dx_t]}{dt}.$$

So, $\hat{\pi}_t = -\delta p_t + \frac{\mathbb{E}_t[dp_t]}{dt}$ (see Section E.3) becomes:

$$\hat{\pi}_t = (D - \delta) p_t \tag{119}$$

and (118) gives

$$Df_t = \chi \hat{\pi}_t. \tag{120}$$

The basic dynamic pricing equation, (76), becomes:

$$0 = -\zeta p_t + \kappa D p_t + f_t. \tag{121}$$

Differentiating once and taking time-t expectations gives:

$$\begin{aligned} 0 &= -\zeta D p_t + \kappa D^2 p_t + D f_t && (122)\\ &= \left[-\zeta D + \kappa D^2 + \chi (D - \delta)\right] p_t && (123)\\ &= H(D) p_t \end{aligned}$$

where

$$H(x) = \kappa x^2 - (\zeta - \chi) x - \chi\delta. \tag{124}$$

The fundamental solutions of equation $H(D) p_t = 0$ are of the form $p_t = Be^{xt}$, with $H(x) = 0$.

There are two roots to $H(x) = 0$, of opposite sign: we call them ρ and $-\phi$, with ρ and ϕ weakly positive:

$$\rho = \frac{\zeta - \chi + \sqrt{\Delta}}{2\kappa}, \qquad \phi = \frac{-\zeta + \chi + \sqrt{\Delta}}{2\kappa}, \qquad \Delta = (\zeta - \chi)^2 + 4\chi\kappa\delta. \tag{125}$$

When $\chi = 0$, $\rho = \frac{\zeta}{\kappa}$ (as in Proposition 5) and $\phi = 0$. We record that ϕ solves:

$$(\zeta + \kappa\phi)\phi = \chi(\phi + \delta) \tag{126}$$

and as the product of the two roots, $-\phi\rho$ is equal to $\frac{-\chi\delta}{\kappa}$ in (124),

$$\phi = \frac{\chi\delta}{\kappa\rho}. \tag{127}$$

This allows us to derive a variety of impulse responses. Calling y_t the process $dy_t = -\phi y_t dt + \sigma dz_t$, let us look for a solution of the form (or "Ansatz"):

$$p_t = A y_t, \qquad f_t = a y_t.$$

Plugging this Ansatz in (118) and examining the σz_t term gives:

$$a = 1.$$

Next, we have $Dy_t = -\phi y_t$, so plugging this in (121) gives: $0 = [-(\zeta + \kappa\phi) A + a] y_t$, i.e.

$$A = \frac{1}{\zeta + \kappa\phi}.$$

This derived Proposition 9 in continuous time.□

One can derive other things. For instance, here's an expression for the price, expressed in discrete time for convenience.

Proposition 11. (Equilibrium price in infinite horizon model, with enriched model of households)*Suppose that* $\Delta f_t = \chi\hat{\pi}_t + \Delta\tilde{f}_t$ *for some arbitrary* $\tilde{f}_t$*. Then, the price at time* t *is:*

$$p_t = \frac{f_{t-1}}{\bar{\zeta}} + \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho}{(1+\rho)^{\tau-t+1}} \left(\frac{\tilde{f}_\tau - \tilde{f}_{t-1}}{\bar{\zeta}} + M^D d_\tau^e \right) \tag{128}$$

with $\bar{\zeta} = \zeta + \kappa\phi$ and $M^D = \frac{\kappa\delta}{\bar{\zeta}}$.

This generalizes Proposition 5. The economics is largely the same, except that ζ is replaced by $\bar{\zeta}$, the expression of ρ changes to (125), and the price impact of an inflow $\Delta\tilde{f}_t$ decays at a rate ϕ.

E.6.2 Discrete time

We use the operator

$$\nabla x_t = \mathbb{E}_t [x_{t+1} - x_t]$$

so that $\nabla x_t \simeq Dx_t \times \Delta t$. So $\hat{\pi}_t = -\delta p_t + \mathbb{E}_t [\Delta p_{t+1}]$ becomes:

$$\hat{\pi}_t = (\nabla - \delta) p_t.$$

Likewise, $\Delta f_t = \chi\hat{\pi}_t + \varepsilon_t$ (see (117)) implies:

$$\nabla f_t = \chi\hat{\pi}_{t+1} = \chi (\nabla - \delta) p_{t+1} = \chi (\nabla - \delta)(1 + \nabla\omega) p_t$$

with $\omega = \Delta t = 1$ in discrete time, and a formal sense that we clarify below, $\omega = 0$ in the continuous time limit.

Then, the basic equation (76) becomes:

$$0 = -(\zeta - \kappa\nabla) p_t + f_t.$$

Premultiplying by ∇ gives:

$$\begin{aligned} 0 &= -\nabla(\zeta - \kappa\nabla) p_t + \nabla f_t \\ &= -\nabla(\zeta - \kappa\nabla) p_t + \chi(\nabla - \delta)(1 + \nabla\omega) p_t \\ &= \tilde{H}(\nabla) p_t \end{aligned}$$

with

$$\tilde{H}(x) = (\kappa + \chi\omega) x^2 - (\zeta - \chi(1 - \delta\omega)) x - \chi\delta. \tag{129}$$

Polynomial $\tilde{H}(x)$ is the discrete-time analogue to the continuous time polynomial $H(x)$ seen above.

Then, we call ρ and $-\phi$ the roots of polynomial $\tilde{H}$.

Now, defining $y_t = (1-\phi)\, y_{t-1} + \varepsilon_t$, we seek solutions of the type:

$$p_t = A y_t, \qquad f_t = a y_t. \tag{130}$$

This implies

$$\hat{\pi}_t = (\nabla - \delta)\, p_t = -(\phi + \delta)\, A y_t.$$

Plugging this in (117) gives:

$$a = -\chi\omega\,(\phi+\delta)\, A + 1.$$

Plugging the Ansatz (130) in (76) gives:

$$0 = -(\zeta + \kappa\phi)\, A + a.$$

Hence, we obtain: $A = \frac{a}{\zeta+\kappa\phi}$, with

$$a = \frac{1}{1 + \chi\omega\frac{\phi+\delta}{\zeta+\kappa\phi}}. \tag{131}$$

Again, formally, we obtain the continuous time limit when $\omega \to 0$.

From discrete to continuous time We draw from Section E.13.1, and denote with bolded symbols the continuous-time version of the parameters, and Δt the calendar value of a time interval. As they are unitless, χ and ζ are the same in discrete and continuous time.

We have $\omega = \Delta t$, as we wrote $\mathbb{E}_t[p_{t+1}] = (1 + \nabla\omega)\, p_t$. Hence, calling $x = \boldsymbol{x}\Delta t$,

$$\begin{aligned} \tilde{H}(x) &= (\kappa + \chi\omega)\, x^2 - (\zeta - \chi(1 - \delta\omega))\, x - \chi\delta \\ \tilde{H}(x)/\Delta t &= (\boldsymbol{\kappa} + \chi\Delta t)\, \boldsymbol{x}^2 - (\zeta - \chi(1 - \delta\Delta t))\, \boldsymbol{x} - \chi\delta \end{aligned}$$

so that indeed, $\lim_{\Delta t\to 0} \frac{\tilde{H}(x)}{\Delta t} = H(\boldsymbol{x})$.

E.6.3 Impact of a trend on dividends

We prove the following.

Proposition 12. *Suppose that $d_t = gt$. Then, if flows follow*

$$\Delta f_t = \chi\hat{\pi}_t + (1-\theta)\, g + c$$

with $f_{-1} = 0$,with $\chi > 0$ and some constant c. Then in the long run, the equity premium is higher, $\hat{\pi}_* = -\frac{c}{\chi}$. and we have $p_t = d_t + p_*$, $f_t = (1-\theta)\, d_t + f_*$, with $p_* = \frac{c}{\chi\delta}$, i.e.

$$p_* = \frac{c}{\chi\delta} = \frac{c}{\kappa\phi\rho} \tag{132}$$

and $f = \zeta p_*$. For finite t, we have

$$p_t = d_t + \left(1 - \frac{\zeta}{\zeta+\kappa\phi}\,(1-\phi)^t\right) p_* \tag{133}$$

so that on impact

$$p_0 = \frac{c}{(\zeta+\kappa\phi)\,\rho} \tag{134}$$

where ρ,$-\phi$ are the of the characteristic polynomial $H(x)$ in (124). The flows are

$$f_t = (1-\theta)\, d_t + f_*\left(1 - (1-\phi)^t\right). \tag{135}$$

We write the rule as a deviation c from the rational flow, which is $\Delta f_t = (1-\theta)\, g$ by Lemma 1. In the baseline case $\Delta f_t = \chi \hat{\pi}_t + \varepsilon_t$, then $c = -(1-\theta)\, g < 0$. Intuitively, if there is a low χ, the "flows don't adjust enough", so that the price is too low, and the equity premium is higher. This is why the intercept p_* is negative.

Also, the long run impact is larger than the short run impact, because the mistakes c "pile up" over time. The speed of convergence is ϕ, which is about 9%. So, for most purposes, the impact p_0 is more important than the long run impact.[94]

Proof. First, we derive the long run, which is simpler. Calling $\hat{\pi}_*$ the steady state deviation of the equity premium from $\bar{\pi}$, we have on average $\Delta f_t = \chi \hat{\pi}_* + (1-\theta)\, g + c$. But Lemma 1 showed that we need $\Delta f_t = (1-\theta)\, g$. So, this implies $\hat{\pi}_* = \frac{-c}{\chi}$. This in turn corresponds to $p_t = p_* + gt$, with $p_* = -\frac{\hat{\pi}_*}{\delta}$.

Next, we derive the finite-time behavior. For simplicity, we use continuous time, and set $g = 0$ for simplicity (the general case is similar). We have $Df_t = \chi \hat{\pi}_t + c$. Insert this in (123) gives

$$H(D)\, p_t + c = 0. \tag{136}$$

The solution is $p_t = p_* + Ae^{-\phi t} + Be^{\rho t}$ for constants A and B. The large t behavior implies $B = 0$. As time $t = 0$, we must have $f_0 = 0$, so

$$0 = -\zeta p_0 + \kappa \frac{\mathbb{E} dp_0}{dt} = -\zeta p_* - (\zeta + \kappa\phi)\, A.$$

This gives $A = -\frac{\zeta}{\zeta + \kappa\phi} p_*$. This implies

$$p_0 = p_* + A = \left(1 - \frac{\zeta}{\zeta + \kappa\phi}\right) p_* = \frac{\kappa\phi}{\zeta + \kappa\phi} \frac{c}{\chi\delta}.$$

We use that $\phi = \frac{\chi\delta}{\kappa\rho}$ from (127), which gives (134).

Finally, as $0 = -\zeta p_t + \kappa D p_t + f_t$, we have

$$f_t = \zeta p_* + (\zeta + \kappa\phi)\, A e^{-\phi t} = \zeta p_* \left(1 - e^{-\phi t}\right).$$

Using our calibration $g = 2\%$, $\theta = 0.85$ and rule (68) with $\chi = 0.23$ we find: $\hat{\pi}_* = \frac{1-\theta}{\chi} g = 1.2\%$, a moderate effect.

We next prove a result that synthesizes and expands on our previous results. □

Proposition 13. *Suppose an economy with i.i.d. dividend growth $\Delta d_t = g + \varepsilon_t^d$, and consumer flows following the semi-behavioral rule:*

$$\Delta f_t = \chi \hat{\pi}_t + (1 - \theta + \gamma)\, \Delta d_t + c + \varepsilon_t^f \tag{137}$$

where disturbances ε_t^d, ε_t^f have mean 0 and no time correlations. The rational case obtains when γ, χ, c, var $\left(\varepsilon_t^f\right)$ are set to 0. Then, in the steady state, the equity premium is $\bar{\pi} + \hat{\pi}$ with $\hat{\pi} = -\frac{\gamma g + c}{\chi}$, and using $p_* = -\frac{\hat{\pi}}{\delta}$, $f_* = \zeta p_*$, and ϕ the mean-reversion of Proposition 9,

$$f_t = (1-\theta)\, d_t + \hat{f}_t + f_*, \qquad p_t = d_t + \frac{\hat{f}_t}{\zeta + \kappa\phi} + p_* \qquad \hat{f}_t = (1-\phi)\, \hat{f}_{t-1} + \varepsilon_t^f + \gamma \varepsilon_t^d. \tag{138}$$

[94] Note that it could be obtained in the case $\chi \to 0$ from Proposition 5, which gives $p_0 = \frac{c}{\zeta\rho}$ (by plugging in $f_\tau = c\tau$).

Here in (137), γ is a "gap", as the rational case would entail $\gamma = 0$. So, the "gap" creates a permanent change in the equity premium (as in Proposition 12). The new part is really the impact of disturbances ε_t^d, ε_t^f: their impact means-reverts at the rate ϕ. Excess flows ε_t^f make the price temporarily too high, and high dividends not immediately compensated by a flow ($\gamma\varepsilon_t^d$) make the price temporarily too low, as in the myopia effect of Proposition 3.

Proof. The terms corresponding to the non-zero trend g and c are exactly as in Proposition 12, using $c' = \gamma g + c$. So, by linearity, we can set $g = c = 0$ and focus on the stochastic terms. In the case where $\varepsilon_t^d = 0$, which is exactly Proposition 9. Then, the case ε_t^d is very similar, as the "mistake" in flows is the sum of the shock ε_t^f and the "excess adjustment" $\gamma\varepsilon_t^d$. □

E.7 Corporate finance in inelastic markets: Complements

We provide complements to Section 6.2.

E.7.1 An increase in buyback financed by a decrease in dividends: Infinite horizon

Here we complete the discussion of the main text, with the infinite horizon case.

We provide a simple thought experiment. Suppose that at time 0 there is a permanent change in the share buyback policy: corporations devote a fraction b of their dividend payout to share buybacks, the rest to dividends (see Boudoukh et al. (2007c) for an empirical analysis). So, the aggregate dividend goes from $\mathcal{D}_t$ to $\mathcal{D}'_t = \mathcal{D}_t(1-b)$,and at each date corporations spend $b\mathcal{D}_t$ on share buybacks. To streamline the computations, we use continuous time.

Buybacks in a frictionless rational model We first consider the rational model.

Proposition 14. *Consider a firm that at time 0 changes its payout policy, and devotes a fraction b of the payout to share buybacks, $1-b$ to dividends (starting from paying only dividends before time 0). Consider a frictionless, rational model. Then, the dividend price ratio falls by a factor $1-b$:*

$$\delta' = (1-b)\,\delta.$$

It goes from $\delta = r + \pi - g$ to $\delta' = (1-b)\,\delta = r + \pi - g'$, where $g' = g + G$ if the new average growth rate of the dividend per share, which is increased by $G = b\delta$. At the same time, the market value of the firm is unchanged, as per Modigliani-Miller.

The surprise is that it changes the D/P ratio by a big amount. The share of dividend as a fraction of the payout has moved from roughly 100% to 50%, so $b = 0.5$. Hence, Proposition 14 implies that the price-dividend ratio went from δ (empirically, about 4%) to half its value (about 2%). This is simply because the growth rate per share has increased by $G = 2\%$, because the number of shares has decreased by the same $G = 2\%$.

Proof. We had $P_tQ_t = \frac{\mathcal{D}_t}{\delta}$. As $b\mathcal{D}_t$ dollars are devoted to purchasing shares each period, the number of shares follows $\dot{Q}_t = -\frac{b\mathcal{D}_t}{P_t} = -bQ_t\delta$, so that the new number of shares is:

$$Q'_t = Q_0 e^{-Gt}, \qquad G = b\delta. \tag{139}$$

Hence, the dividend per share in the new regime is $D'_t = \frac{(1-b)\mathcal{D}_t}{Q'_t} = \frac{(1-b)\mathcal{D}_t}{Q_0 e^{-Gt}}$ i.e. as $D_t = \frac{\mathcal{D}_t}{Q_0}$,

$$D'_t = (1-b)\, e^{Gt} D_t. \tag{140}$$

Because the value of the firm is constant, $P'_t Q'_t = P_t Q_0$, we have

$$P'_t = P_t e^{Gt}. \tag{141}$$

This implies that the new D/P ratio is

$$\frac{D'_t}{P'_t} = \frac{D_t}{P_t}(1-b). \tag{142}$$

This is of course consistent with the Gordon formula: as $\delta = r_f + \pi - g$, as under the new regime the dividend per share grows at a rate $g + G$, we have

$$\delta' = r_f + \pi - g - G = \delta - b\delta = (1-b)\,\delta,$$

using (139) in the last equation. □

Buybacks in an inelastic model We next study the situation in an inelastic model.

Proposition 15. (Impact of share buybacks in the infinite horizon model) *In the inelastic model, suppose that a change in policy on a scale b is announced, and should last forever. Then, with $c = \left(\mu^D - \mu^G\right)\theta\delta b$, the firm value increases on impact by $v_0 = \frac{c}{(\zeta+\kappa\phi)\rho}$ and in the long run by $v_* = \frac{c}{\kappa\phi\rho}$. In between, the change in firm value is $v_t = v_* + (v_0 - v_*)\,(1-\phi)^t$.*

Hence, the economics is similar to the simple model of Proposition 8.

Proof. As in the rational case, growth rate in the number of shares is $-G$ with $G = \delta b$, $d_t = -b + Gt$. As at each period the dividends are lower by b, and they represented a fraction $\delta\theta$ of the fund's value, $f^{C\to M} = -\int_0^t \delta b\theta dt = -\theta G t$, and $q^S = -Gt$. So, if households do not change their flows ($\mu^D = \mu^G = 0$), by Proposition 4, we have $p_t = Gt$. Hence, the value of the firm is unchanged.

Now, consider the case where households do change their flows: as each period the dividends are lower by b, and they represented a fraction $\delta\theta$ of the fund's value, and capital gains are increased by G, we have

$$\Delta f^{h\to M} = -\mu^D(-b\delta\theta) - \mu^G G\theta = \left(\mu^D - \mu^G\right)\theta G.$$

Equilibrium is

$$-\zeta p_t + \kappa \mathbb{E}\Delta p_{t+1} + f^{C\to M} + f^{h\to M} = q^S$$

i.e.

$$-\zeta p_t + \kappa \mathbb{E}\Delta p_{t+1} + f_t = 0$$

with

$$f = -q^S + f^{C\to M} + f^{h\to M} = \left(\mu^D - \mu^G\right)\theta G t + (1-\theta)\,Gt.$$

So, we are in the situation of Proposition 12, with $g = G = \delta b$, $c = \left(\mu^D - \mu^G\right)\theta G$. Hence, the value of the firm increases on impact by:$p_0 = v_0 = \frac{c}{(\zeta+\kappa\phi)\rho}$ and in the long run by $v_* = p_* = \frac{c}{\kappa\phi\rho}$. □

Tentative calibration We proceed as above, with $\mu^D = 0.5$, $\mu^G = 0.03$. Under this tentative calibration, a buyback of 1% of the market increase the market capitalization by 0.6% in the short run, and 1.6% in the long run.

E.7.2 An increase in buyback financed by increased debt rather than lower contemporaneous dividends.

Consider an increase in buybacks that is not compensated by a contemporaneous decrease in dividends — so that the total payout is increased, by a factor equal to b times the initial market value.

Two period model. The buyback of B dollars decreases the number of shares by $\frac{B}{P_0}$, the future aggregate dividend by BR, where R is the gross interest rate. The time-1 dividend is $\mathcal{D}_1' = \mathcal{D}_1 - BR$, the present value of $\mathcal{D}_1'$ falls by a fraction b. As the number of shares also falls by b, the present value of the time-1 dividend per share remains constant:

$$q^S = -b, \qquad d = 0.$$

In a frictionless model, this buyback does not change the current price per share, and does not change the time-0 return r

$$\text{Frictionless model: } p = 0, \qquad v = -b, \qquad r = 0.$$

In an inelastic model, now $q^D = -\zeta p + f^h = q^S = -b$, so

$$p = \frac{b + f^h}{\zeta}, \qquad v = p - b, \qquad r = p.$$

Hence, the aggregate values of equities increases, and the time-0 return r is positive, unless it's compensated by a flow $f^h = -b$. Using the marginal propensities in Section 6.2, we have $f^h = -\mu^G \theta p$, so that in total:

$$p = \frac{b}{\zeta + \theta \mu^G}. \tag{143}$$

Infinite horizon. We suppose that the debt will be repaid very far in the future (at date $T \to \infty$.) Then, the economics is as in the two-period model.[95]

E.8 General equilibrium in a two-period economy

We detail a two-period endowment economy. This clarifies a number of things, such as the link between consumption and cash-flows, and the determination of the interest rate.

There are a representative consumer and a representative firm in an endowment economy. The consumer has utility

$$U(C_0, C_1) = u(C_0) + \beta u(C_1). \tag{144}$$

In equilibrium all consumption at time $t \in \{0, 1\}$ comes from a non-storable endowment Y_t. The time-0 endowment Y_0 has been given to the consumer as a time-0 dividend. Output Y_1 is stochastic, with a strictly positive lower bound.

[95] In a rational model, we still have $p_t = 0$, $v_t = -b$. In an inelastic model, as $\rho > \delta$,we have $d_t = 0$, $q_t^S = -b$, so $p_t = \frac{b+f^h}{\zeta}$ (for all dates $t \ll T$) also. Hence the expression is as in the two-period model.

The fruit from the tree goes to the representative firm, which pays households as debt repayment and stock dividend, Y_1.

At time 0, the government issues a bond in quantity Q^B, and gives it to the citizens. It is backed by taxes $R_f Q^B$ collected at time 1. This does not affect aggregate consumption, but it clarifies the origins of bonds. One could also just set $Q^B = 0$. Then, the pure bond fund will hold a negative quantity of bonds.

We call $R_1^{\mathcal{E}} = \frac{Y_1}{P_0}$ the gross return on equities. If the consumer sells I of bonds to invest them in equities, and saves an additional s dollars (so that $C_0 = Y_0 - s$), time-1 wealth is:

$$W_1 = \left(P_0^{\mathcal{E}} Q^{\mathcal{E}} + I\right) R_1^{\mathcal{E}} + \left(Q^B - I + s\right) R_f - Q^B R_f.$$

The first term is the proceeds from equities, the second term is the proceeds from the pure bond fund, and the last term the taxes paid to the government. This simplifies to:

$$W_1(I, s) = \left(P_0^{\mathcal{E}} Q^{\mathcal{E}} + I\right) R_1^{\mathcal{E}} + (-I + s) R_f. \tag{145}$$

A traditional, rational benchmark Let us examine what would happen in the traditional, rational benchmark. The consumer's problem would be

$$\max_{s,I} u(Y_0 - s) + \mathbb{E}\left[u\left(W_1(I, s)\right)\right].$$

In equilibrium, we should have $s = I = 0$. Hence, Lucas (1978) pricing would hold. We would have a rational pricing kernel:

$$\mathcal{M} = \beta \frac{u'(Y_1)}{u'(Y_0)}. \tag{146}$$

The interest rate R_f would satisfy

$$R_f \mathbb{E}[\mathcal{M}] = 1, \tag{147}$$

and the time-0 price of equities would be:

$$P_0^* = \mathbb{E}[\mathcal{M} D_1]. \tag{148}$$

But we wish to propose an alternative to that Lucas (1978) approach.

As in the simple model of Section 3, in addition to a pure bond vehicle there is a balanced allocation fund that invests in a mix of bonds and equities. The market value of the mixed fund is $W_t^M = Q_{Mt}^B + Q_{Mt}^{\mathcal{E}} P_t$ at dates $t = 0^-$ and $t = 0$. Its value at time $t = 1$ is $W_1^M = Q_{M0}^B R_f + Q_{M0}^{\mathcal{E}} D_1$.

We assume that we are in equilibrium at time 0^-, before the flow and dividend news. By this we mean that households have allocated their holdings to the bond and mixed fund in the right proportion, so that equity values are correct, $P_{0^-} = P_0^*$ as in (148). As the mixed fund allocates a fraction θ in equity, and $1 - \theta$ in bonds, it manages a wealth $W_{M,0^-} = \frac{1}{\theta} Q^{\mathcal{E}} P_0^*$, and a quantity of bonds $Q_{M,0^-}^B = \frac{1-\theta}{\theta} Q^{\mathcal{E}} P_0^*$. The pure bond fund holds the remaining bonds, $Q^B - Q_{M,0^-}^B$.

The household is endowed with all shares of the pure bond fund and mixed fund. He may wish to increase his savings by s; and he may wish to invest an additional ΔF dollars in the mixed fund, taken from selling ΔF from the pure bond fund. So, the consumptions are:

$$C_0 = Y_0 - s, \qquad C_1 = Y_1 + s R_f + \left(\frac{W_1^M}{W_0^M} - R_f\right) \Delta F,$$

which defines $C_t(s, \Delta F)$ and the expected utility

$$U(s, \Delta F) = u(C_0(s, \Delta F)) + \beta \mathbb{E}[u(C_1(s, \Delta F))].$$

If the household were rational, its plan would be to maximize U subject to the above budget constraint: $\max_{s,\Delta F} U(s, \Delta F)$.

Flows and behavioral assumptions We model households as rational in consumption, but behavioral in portfolio choice. More formally, we decompose the household as a consumer, who can only trade the pure bond fund (so, chooses s) and decide on consumption (given the money in the pure bond fund), and an investor, who trades the different funds (so chooses ΔF), but cannot decide on consumption.

The consumer is rational, and she "sees through" all the cash-flows. Her plan is $\max_s U(s, \Delta F)$. She sees that in equilibrium she will consume the endowment Y_t at time t. As she trades the bond, she enforces the Euler equation, so that (147) holds with pricing kernel (146).

Suppose that at time 0 there is a shock to the expectation of the future output Y_1 and a flow shock ΔF. To concentrate on the essentials, we assume that this shock does not change the frictionless interest rate R_f. Then, the economy is as described in Section 3. In particular, the price changes as in that simple model.

E.9 Details of the household's problem in general equilibrium

This section provides some extra details to the household's problem of Section 5.1.

We call t^- the beginning of period values, evaluating all at the time t price P_t. The mixed fund gives a dividend $D_t^M = Q^{\mathcal{E}} Y_t + r_{f,t-1} B_{t-1}^M$, so that its cum-dividend value is $W_{t^-}^M = Q^{\mathcal{E}} P_t + B_{t-1}^M + D_t^M$, and the return is $R_t^M = \frac{W_{t^-}^M}{W_{t-1}^M}$.

The mixed fund has issued N_{t-1} shares, of which N_{t-1}^h are owned by household h. The value of a share in the mixed fund is $v_t^M = \frac{Q^{\mathcal{E}} P_t + B_{t-1}^M}{N_{t-1}}$. So, the beginning of period wealth of the household is:

$$W_{t^-}^h = \frac{N_{t-1}^h}{N_{t-1}} W_{t^-}^M + B_{t-1}^h R_{f,t-1}. \tag{149}$$

Suppose the households flow ΔF_t^h into the mixed fund, while the rest of the economy flows ΔF_t (in equilibrium, the two values are the same). Then, the number of shares owned by the household and in the fund are:[96] $N_t^h = N_{t-1}^h + \frac{\Delta F_t^h}{v_t^M}$ and $N_t = N_{t-1} + \frac{\Delta F_t}{v_t^M}$. The household holds B_t^h in the pure bond fund:

$$B_t^h = B_{t-1}^h R_{f,t-1} + \frac{N_{t-1}^h}{N_{t-1}} D_t^M - C_t - \Delta F_t^h,$$

i.e. the proceeds from the pure bond fund, the dividend of the mixed fund, minus consumption, minus the flow.

The household's problem, in its rational form, is:

$$V(W_{t^-}, Z_t) = \max_{C_t, B_t^h} u(C_t) + \beta \mathbb{E}\left[V\left(W_{t+1^-}^h, Z_{t+1}\right)\right].$$

[96] We also have $v_t^M = \frac{W_t^M}{N_t} = \frac{Q^{\mathcal{E}} P_t + B_{t-1}^M}{N_{t-1}}$, $B_t^M = B_{t-1}^M + \Delta F_t$. Flows change the number of shares issued by the fund, but not (controlling for stock prices) the value of each fund share.

This problem defines a consumption, and also desired holdings in the pure bond fund (hence, a flow out of the bond fond).

E.10 Pricing kernel consistent with flow-based pricing: Complements

Much of asset pricing uses pricing kernels, or stochastic discount factors (SDFs). We show how to express the economics of flows in inelastic markets in the language of pricing kernels. To do so, we outline a simple general method to complete a "default" pricing kernel so that it reflects the impact of flows on asset prices.

Pricing kernel completion: How to adjust a default pricing kernel to reflect the impact of flows on asset prices For simplicity, we omit the time subscripts.

Default pricing kernel. We allow for a "default pricing kernel", which prices bonds at the equilibrium interest rate R_f. The simplest is the "risk-free" default pricing kernel: $\mathcal{M}^d = \frac{1}{R_f}$.[97]

From the default pricing kernel to the actual pricing kernel. The default pricing kernel $\mathcal{M}^d$ will not price assets correctly, as it does not react to flows. We propose a method of "pricing kernel completion" that will augment the pricing kernel so that it correctly prices all assets. We posit the existence of a very small mass ε (which we will take to be infinitesimal, so that it won't impact prices) of "agile optimizers," who start with zero financial wealth and whose objective function is:

$$\max_Q \mathbb{E}\left[-\mathcal{M}^d e^{-Q'R}\right], \tag{150}$$

where R is the vector of excess returns at time 1.[98] That is, they maximize (over a vector Q of holdings over all assets) their expected return $R = \frac{P_1+D_1}{P_0} - R^f$, starting from zero wealth, but this is their expected return "under the risk-neutral probability" generated by $\mathcal{M}^d$.[99] Hence we have $\mathbb{E}\left[\mathcal{M}^d e^{-Q'R} R\right] = 0$. So, the following $\mathcal{M}$ is a pricing kernel:

$$\mathcal{M} = \mathcal{M}^d e^{-Q'R+\xi}, \tag{151}$$

where the constant ξ ensures that the risk-free rate is correctly priced ($\mathbb{E}[\mathcal{M}] = \mathbb{E}\left[\mathcal{M}^d\right]$, so $\xi = \ln \frac{\mathbb{E}\left[\mathcal{M}^d\right]}{\mathbb{E}\left[\mathcal{M}^d e^{-Q'R}\right]}$).

We call this the "completed" pricing kernel. Note that other SDFs could also work (as is generic in incomplete markets), but the one given in (151) is the unique SDF coming from the "pricing kernel completion" procedure. We treated here the simplest case, with just one risky asset, and the simplest default pricing kernel $\mathcal{M}^{d,R_f} = \frac{1}{R_f}$.

Flow-based SDF for the two-period model Let us revisit the two-period model of Section 3.1. The excess equity premium is $\hat{\pi} = \delta(d-p)$ with p given in (10), so that, with $f = (1-\theta)d + \tilde{f}$, the total equity premium is: $\pi = \bar{\pi} - \delta\frac{\tilde{f}}{\zeta}$. So, the completed pricing kernel is:

$$\mathcal{M} = \exp\left(-r_f - \pi\frac{\varepsilon^D}{\sigma_d^2} + \xi\right), \qquad \pi = \bar{\pi} - \delta\frac{\tilde{f}}{\zeta}, \tag{152}$$

[97] In the spirit of maintaining a continuity with the heritage of Lucas (1978), we can also consider a "consumption CAPM" default pricing kernel: $\mathcal{M}^{d,C} = \beta\frac{u'(C_1)}{u'(C_0)}$. We develop this in Section E.10.1.

[98] The implicit risk aversion of 1 is just a normalization.

[99] They start with zero capital at each period, and rebate their profits and losses to the representative household.

with $\xi = -\frac{\pi^2}{2\sigma_d^2}$ if ε^D is Gaussian. This SDF prices correctly stocks and bonds.

This gives the "flow-based" completed pricing kernel, which is an alternative to the consumption-based kernel of Lucas (1978). The core economics is in how flows affect prices, and the pricing kernel (152) just reflects that. If there is a flow f, that modifies the price P according to (10), and the pricing kernel $\mathcal{M}$, in such a way that $P = \bar{P}(1+p) = \mathbb{E}[\mathcal{M}D]$ holds. The pricing kernel is in a sense a symptom rather than a cause in that market.

Flow-based SDF for the infinite-horizon model Section 5.3 developed the SDF for the infinite-horizon model process for flows in (21)-(22), something also delivered by our general equilibrium model of Section 7. A justification is that we assume that "dividend strips" are also traded. By the above procedure we obtain the pricing kernel for each date. In the construction, then dividend strips have an equity premium π_t independent of maturity. So, the maximum Sharpe ratio is achieved via a one-period dividend strip.

Formally, one obtains the price of any asset, once we have a SDF. However, one can reasonably hope to obtain a correct price only when the novel asset is in very small quantity, as the agile optimizers, which form a very small group, will be able to absorb it. When there are substantially different asset classes, one needs to think about flows in those different classes — they will affect prices, and hence the SDF, along the lines we just saw. We next show how easy it is to generalize the model to several asset classes.

.

E.10.1 More general cases to get a pricing kernel

Here we expand Section 5.3 to multiple risky assets and a consumption-based default SDF.

A Gaussian example To be clear let us work out a basic example. We suppose that returns and consumption are Gaussian:

$$\frac{C_1}{C_0} = e^{g_c + \sigma_c \varepsilon^c - \frac{1}{2}\sigma_c^2}, \tag{153}$$

with ε_t^c a standard Gaussian random variable. Consider the consumption pricing kernel, which is:

$$\mathcal{M}^{d,C} \equiv e^{\mathcal{M}^{d,C}} = \beta \left(\frac{C_1}{C_0}\right)^{-\gamma} = e^{-r_f - \gamma\sigma_c\varepsilon^c - \frac{1}{2}\gamma^2\sigma_c^2}$$

for the risk-free rate $r_f = -\ln\beta + \gamma g_c - \frac{1}{2}\gamma(1+\gamma)\sigma_c^2$.

We next consider the agile optimizers' problem, going back to a general default pricing kernel $\mathcal{M}^d$ (that might be $\mathcal{M}^{d,R_f}$ or $\mathcal{M}^{d,C}$). We recall that for two jointly Gaussian variables X, Y:

$$\frac{\mathbb{E}\left[e^X Y\right]}{\mathbb{E}\left[e^X\right]} = \mathbb{E}[Y] + cov(X, Y). \tag{154}$$

For instance, the anomalous excess equity premium is

$$\mathbb{E}^{\mathcal{M}^d}[R] := \frac{\mathbb{E}\left[\mathcal{M}^d R\right]}{\mathbb{E}\left[\mathcal{M}^d\right]} = \mathbb{E}[R] + cov\left(\mathcal{M}^d, R\right), \tag{155}$$

which is the expected excess return of R that is not explained by the default pricing kernel: indeed, if the pricing kernel $\mathcal{M}^d$ correctly priced R, we'd have $\mathbb{E}^{\mathcal{M}^d}[R] = 0$. Put another way, those are the excess returns above and beyond what is warranted by the default pricing kernel.

The FOC of (150) is $\mathbb{E}\left[\mathcal{M}^d e^{-Q'R} R\right] = 0$, so that using (154), with $V_R = cov(R, R)$ the variance-covariance matrix of returns,

$$\mathbb{E}[R] + cov\left(\mathcal{M}^d, R\right) - V_R Q = 0,$$

where $-cov\left(\mathcal{M}^d, R\right)$ is the equity premium warranted by the default pricing kernel. The optimal portfolio of agile optimizers is $Q = V_R^{-1}\mathbb{E}^{\mathcal{M}^d}[R]$ and their return is a form of "tangency portfolio" return:

$$R^\tau = Q'R = \mathbb{E}^{\mathcal{M}^d}[R]' V_R^{-1} R, \tag{156}$$

which depends on the "anomalous" excess returns $\mathbb{E}^{\mathcal{M}^d}[R]$. Their "excess Sharpe ratio" is

$$\mathcal{S} = \frac{\mathbb{E}^{\mathcal{M}^d}[R^\tau]}{\sigma_{R^\tau}}, \tag{157}$$

which is the Sharpe ratio they get in excess of the average returns warranted by the default pricing kernel. Given that $\mathbb{E}^{\mathcal{M}^d}[R^\tau] = \mathbb{E}^{\mathcal{M}^d}[R]' V_R^{-1}\mathbb{E}^{\mathcal{M}^d}[R] = \sigma^2_{R^\tau}$, we have $\mathcal{S} = \sigma_{R^\tau}$. Hence, the SDF is their marginal utility (up to a proportional factor that is pinned down by the risk-free rate), which is

$$\mathcal{M} = \mathcal{M}^d e^{-\mathcal{S}\frac{R^\tau - \mathbb{E}^{\mathcal{M}^d}[R^\tau]}{\sigma_{R^\tau}} - \frac{1}{2}\mathcal{S}^2}. \tag{158}$$

This SDF $\mathcal{M}$ prices all assets correctly: $P_a = \mathbb{E}[\mathcal{M} D_a]$ for all assets.

E.11 On the link between the Kyle lambda and the market inelasticity

E.11.1 Theory: Kyle's lambda versus inelasticity

Suppose that within a certain time window, there is an "order flow" (realized signed trades), with volume Δf_t expressed as a fraction of the market capitalization. A typical microstructure regression is, as in Hasbrouck (2007):

$$p_t - p_{t-1} = \lambda\left(\Delta f_t - \mathbb{E}_{t-1}[\Delta f_t]\right) \tag{159}$$

where λ is the so-called "Kyle lambda", from Kyle (1985). We analyze what that regression would estimate in our model.

We suppose that our model holds, and that there is completely symmetric information about fundamentals – so, we remove the informational ingredient of Kyle. Still, trades will move prices – because of inelasticity. We clarify this here. As we mentioned above, a very important difference is that in Kyle flows do not change the equity premium on average, whereas in our model, positive inflows lower the equity premium.

To analyze what happens in our model, $\approx$ in the order flow (like Madhavan et al. (1997), Lillo et al. (2005)and Bouchaud et al. (2018)):

$$\Delta f_t = (1 - \phi_g)\,\Delta f_{t-1} + \varepsilon_t, \tag{160}$$

where ε_t is i.i.d. So, an innovation ε_t creates an innovation to the eventual cumulative flow:[100]

$$\lim_{h\to\infty} \mathbb{E}_t \left[f_{t+h} - f_{t-1} \right] = K\varepsilon_t, \qquad K = \frac{1}{\phi_g}.$$

For instance, if a large desired trade ("meta-order") is on average "sliced" into 15 trades, executed slowly over time, then $K = 15$. Likewise, if a fast fund trades, and is followed on average by similar or "copycat" metaorders by two other funds, then $K = 3$.[101] The two forces combine: if a fund splits its metaorder in five trades, and it is followed by two more similar funds doing a similar trade (also splitting their trade into five chunks), then $K = 5 \times 3 = 15$, the product of the number of "similar" funds (3 in this example), and the number of "chunks" in which they split their trade (5 in this example).

In our model, the total price impact is, in the limit of small time intervals,[102]

$$\Delta p_t = \frac{K}{\zeta}\varepsilon_t = \frac{K}{\zeta}\left(\Delta f_t - \mathbb{E}_{t-1}\left[\Delta f_t \right] \right).$$

Hence, an econometrician estimating (159), will find:

$$\lambda = \frac{K}{\zeta}. \tag{161}$$

This means that, for the aggregate market, the Kyle lambda is the inelasticity $\frac{1}{\zeta}$ times the persistence parameter K associated with the positive autocorrelation of the order flow.

Most empirical work in microstructure is done at the level of one asset, so that the λ they estimate is

$$\lambda = \frac{K}{\zeta^{\perp}}, \tag{162}$$

where $\lambda^{\perp}$ is the micro-elasticity of Section E.4.

E.11.2 Empirical values from the microstructure literature

Frazzini et al. (2018) find that buying 2.5% of the daily volume of a stock creates a permanent price impact $\Delta p = 15$bp (indeed, it creates a total price impact of 18bp, of which 85% is permanent, see their Figures 2 and 6). Using an annual turnover of 100%, and 250 trading days per year, this

[100] Indeed, ε_t creates an innovation to the cumulative flow f_{t+h} equal to

$$\mathbb{E}_t \left[f_{t+h} - f_{t-1} \right] = \mathbb{E}_t \left[\Delta f_t + \Delta f_{t+1} + \cdots + \Delta f_{t+h} \right] = \varepsilon_t \left(1 + (1-\phi_g) + \cdots + (1-\phi_g)^h \right) = \frac{\varepsilon_t}{\phi_g}\left(1 - (1-\phi_g)^{h+1} \right).$$

[101] More generally (such as in models with multiple time scales, or some form of long memory, see Bouchaud et al. (2018)), K is the "expected value of related orders, given the past". So, the estimation of K is a bit delicate, and not simply the inverse of the speed of mean-reversion of orders. Formally, with $\varepsilon_t \coloneqq \Delta f_t - \mathbb{E}_{t-1}\left[\Delta f_t\right]$, $K = \frac{\partial}{\partial \varepsilon_t} \lim_{h\to\infty} \mathbb{E}_t \left[f_{t+h} - f_{t-1} | \varepsilon_t \right]$.

[102] Away from the limit of small time intervals, the calculation is the following. The price is $p_t = A f_{t-1} + B\Delta f_t$ for two coefficients A, B to determine. Calculations based on Proposition 5 (plug in that expression in (15) with $q_t^D = 0$) show: $A = \frac{1}{\zeta}$ and $B = \frac{1+\frac{\kappa}{\zeta}}{\zeta + \kappa\phi_g}$. So, in the limit of small time intervals (as in Section E.6.2), with $\kappa = \boldsymbol{\kappa}/\Delta t \to \infty$, and ζ constant (as $\delta = \boldsymbol{\delta}\Delta t$, $\delta\kappa$ is constant as $\Delta t \to 0$), we get $B \to \frac{1}{\zeta\phi_g} = \frac{K}{\zeta}$, i.e. $\Delta p_t = \frac{K}{\zeta}\varepsilon_t$.

means that buying a fraction $\Delta q = 2.5\% \times \frac{1}{250} = 1\text{bp}$ of the stock creates a 15bp price impact. Hence, their Kyle lambda is[103]

$$\lambda = \frac{\Delta p}{\Delta q} = \frac{15\text{bp}}{1\text{bp}} \simeq 15.$$

Hence, the prima facie "microstructure" price impact is $\lambda \simeq 15$.[104] This can be compared with our own $M \simeq 5$. However, in terms of our model, their λ reflects the micro-elasticity rather than the macro-elasticity: it is $\lambda = \frac{K}{\zeta^{\perp}}$. As we calibrate $\zeta^{\perp} \simeq 1$, this leads to $K \simeq 15$. This estimate has the interpretation, in inelastic markets with a micro elasticity of 1, that a large market-wide desired trade ("metaorder") is on average split into 15 smaller trades executed over time, by one or several institutions collectively (for example, by three funds pursuing a similar strategy, each splitting their desired position change into five smaller trades).

This factor $K > 10$ may seen surprisingly large, but it is consistent with microstructure data. Bouchaud et al. (2018) report a positive autocorrelation of the decay in the signed of trades $\varepsilon_t = sign\left(\Delta f_t\right)$, qualitatively consistent with the above model. Importantly, it is also roughly quantitatively consistent too. The empirical correlation between the signs of trades, $c\left(h\right) = corr\left(\varepsilon_t, \varepsilon_{t+h}\right)$, is approximately $c\left(h\right) \simeq \frac{0.25}{h^{1/2}}$ for $h \in [1, 10^3]$, which leads to $K = 1 + \sum_{h=1}^{10^3} c\left(h\right) = 16$.[105] This means that a buy trade today announces 15 more buy trades in the future – a large empirical autocorrelation of market orders. We explain this, in this section, by order splitting and copycat trades (which is also Bouchaud et al. (2018)'s interpretation – here we also relate it to the micro elasticity $\zeta^{\perp}$ of the market, by (162)). This gives, we think, a potentially satisfactory unification of the very high impact measured impact of the microstructure literature, and the more moderate impact measured in inelastic markets.

One lesson is that the market microstructure literature finds price impacts that are larger than the ones we find (with a price impact multiplier over 15), which may help dispel some feeling that our estimates are too large. By estimating things at a low frequency, and using a model taking into account the autocorrelation of the order flow, we can structurally relate their price impact estimates to the market micro-inelasticity (since most of the microstructure literature is about the micro elasticity, not the macro elasticity).

E.12 Short-term versus long-term elasticity when funds are inertial

The basic model describes price impacts and quantity adjustments assuming no inertia in funds' reactions. Here we study what happens if funds react with some inertia: this creates additional transitory dynamics.

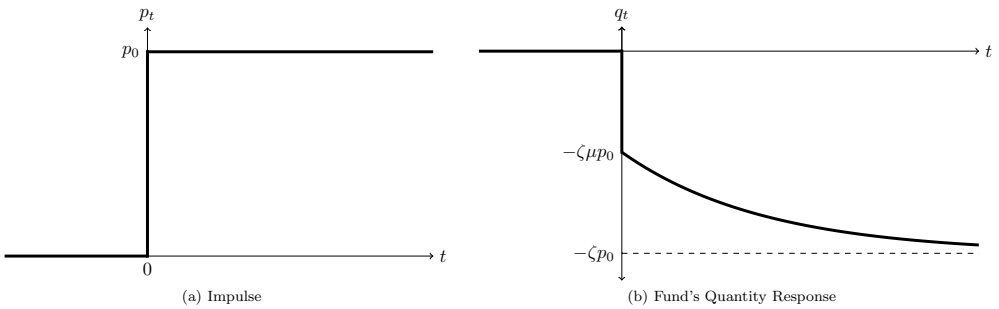
We consider the case of a homogeneous type of fund, trading only the aggregate stock and a risk-free short-term bond. Total demand q_t can change because of the inflow f_t and via an "active"

[103] In practice, the measured price impact is not linear, and indeed looks more concave, perhaps like a square root, which may be due to slower trading of large orders (Torre and Ferrari (1998); Gabaix et al. (2003, 2006); Bouchaud et al. (2018)). We think that this elaboration is beyond the scope of this appendix.

[104] Frazzini et al. (2018) also explore the Trades And Quotes (TAQ) data, and find a price impact about 2.5 times bigger (see their Figure 7). This would would then lead to $\lambda \simeq 37$. In our discussion we use their baseline estimate, which is instead constructed using trading data from AQR, a large institutional asset manager.

[105] See their Figure 10.1. This model has a power law decay rather than an exponential decay, because it is a mixture of several exponential decay. Also, a limitation is that Bouchaud et al. (2018) study the sign of flows, whereas our model would like the signed traded, including their size.

Figure E.10: This figure shows the quantity adjustment of an inert fund after a change in the aggregate stock price. It illustrates Proposition 16. If investors are inertial, there is a gradual adjustment of the quantity over time. When there is no inertia, $\mu = 1$ and the quantity adjustment is instantaneous.



(a) Impulse (b) Fund's Quantity Response

demand q_t^a:

$$q_t = q_t^a + f_t.$$

We model the actual active demand with inertia as:

$$\Delta q_t^a = \mu \Delta q_t^{a,\nu} + \phi \Delta t \left(q_{t-1}^{a,v} - q_{t-1}^a \right), \tag{163}$$

where $q_t^{a,v}$ is the "virtual active demand" – the one of a non-inertial fund:

$$q_t^{a,v} = -\left(1-\theta\right) p_t + \kappa \hat{\pi}_t + \nu_t = -\zeta p_t + \kappa \left(\mathbb{E} p_{t+1} - p_t \right) + \kappa \delta d_t + \nu_t, \tag{164}$$

with $\mu \in [0,1]$ and $\phi \geq 0$. A very frictional investor has $\mu = 0$ and $\phi \geq 0$ low. A frictionless investor has $\mu = 1$. More frictional investors have lower $\mu \geq 0$ and lower $\phi \geq 0$. The adjustment to flows f_t is instantaneous for simplicity, and as it does not require a "strategic" decision by the fund, which simply rescales its investment after an inflow.

We derive quantity adjustments (the proof is by plug and verify).

Proposition 16. (Short-run versus long-run elasticity of demand) *Suppose a fund that exhibit inertia. Then, its short-run elasticity of demand is $\mu\zeta$, and its long-run elasticity of demand is ζ. More precisely, suppose that the log price of equities jumps by p_0 at time 0, i.e. $p_t = 1_{t \geq 0} p_0$. Then, at $t \geq 0$, the fund's demand change is:*

$$q_t = -\zeta \left(1 + \left(\mu - 1 \right) \left(1 - \phi \right)^t \right) p_0, \tag{165}$$

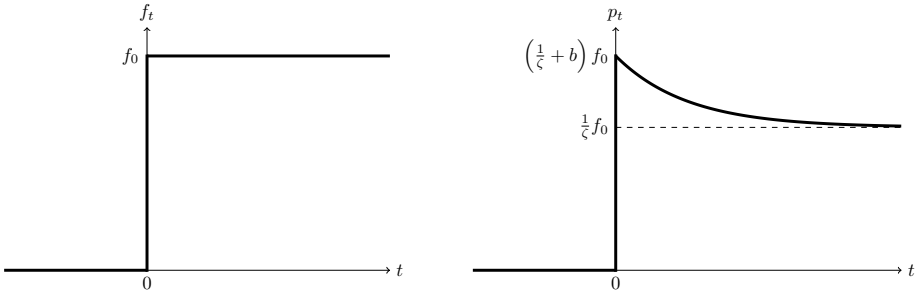
while its virtual demand is $q_t^\nu = -\zeta p_0$.

Figure E.10 illustrates the dynamics. The long run demand is $q_\infty^\nu = -\zeta p_0$, but the impact at 0 is only μ times that, $q_0^\nu = -\zeta \mu p_0$. In between there is an exponential relaxation at rate ϕ.

The next proposition derives the price impact of a flow. [106]

[106] Section E.12 contains the proof, and has complements: it generalizes the price as present value formula (16) and the price impact with inertia formula (166) to the case of flows happening at any date. We also assume $\Phi < 1$, which is automatically true in the limit of small time intervals.

Figure E.11: This figure shows the price dynamics caused by an unanticipated time-0 demand shock, when investors are inertial. It illustrates Proposition 17. The permanent demand shock f_t creates a permanent price change $p_\infty = \frac{f_0}{\zeta}$. If investors are inertial, there is a small extra bump b on impact, that decays exponentially over time. When investor are not inert, $b = 0$ and the price immediately jumps to its permanent value p_∞.



(a) Impulse (b) Price Response

Proposition 17. (Price impact of an inflow when funds are inertial) *When funds exhibit inertia, the price impact of a permanent, unanticipated inflow f_0 at time 0 is (for $t \geq 0$),*

$$p_t = M_t f_0, \qquad M_t = \frac{1}{\zeta} + b\left(1-\Phi\right)^t \tag{166}$$

where $b = \frac{1-\mu}{\mu(\zeta+\kappa\Phi)}$ and $\Phi = \frac{\phi}{\mu}$. So, the short run price impact is $M_0 = \frac{1}{\zeta} + b$, while the long run impact is $M_\infty = \frac{1}{\zeta}$.

Proof. We conjecture a solution of the type (166). We normalize $f_0 = -1$. Plugging this in (164) gives

$$q_t^{a,v} = \left(1 + b\zeta\left(1-\Phi\right)^t\right) + b\kappa\Phi\left(1-\Phi\right)^t = 1 + b\left(\zeta+\kappa\Phi\right)\left(1-\Phi\right)^t = 1 + c\left(1-\Phi\right)^t.$$

For $t \geq 0$, equilibrium imposes $q_t^a + f_0 = 0$, i.e. $q_t^a = 1$. So (163) gives, for $t > 0$

$$0 = \Delta q_t^a = \mu \Delta q_t^{a,\nu} + \phi\left(q_{t-1}^{a,v} - q_{t-1}\right) = c\left(-\Phi\mu + \phi\right)\left(1-\Phi\right)^{t-1},$$

which leads to $\Phi = \frac{\phi}{\mu}$. At time 0, (163) gives

$$q_0^a = \mu q_0^{a,v} = \mu\left(1 + b\left(\zeta + \kappa\Phi\right)\right).$$

As $q_0^a = 1$, this gives $b = \frac{1-\mu}{\mu(\zeta+\kappa\Phi)}$. □

Figure E.11 illustrates the dynamics of (166). An unanticipated, permanent inflow f_0 at time 0 has an immediate price impact $\left(\frac{1}{\zeta} + b\right) f_0$ that is bigger than the long-run price impact $\frac{f_0}{\zeta}$. The initial "excess reaction" bf_0 dies down at the exponential rate Φ. When funds are not inert, $b = 0$.

For $\mu < 1$, we have $\Phi > \phi$: surprisingly, the speed of price dynamics Φ is faster than the fund-level speed of adjustment of quantities ϕ. This is because the movements of the equity premium creates an incentive for adjustment beyond the "mechanical" speed ϕ.[107]

[107] Imagine that the impulse is a positive inflow, which increases the price. First, the "active" part of the fund strategy wants to sell shares, as the price is high and the equity premium low. But a low "instantaneous share" μ creates a high initial price jump, so a very negative expected return, speeding up the selling of shares: hence, the smaller the μ, the greater the price jump p_0, and the faster the price adjustment Φ.

Calibration We discuss the model calibration.[108] For the fund-level inertia we take $\phi = 1/\text{year}$, so that the half-life is about 0.7 years. We also take the instantaneous sensitivity to events to be $\mu = 0.5$, where the calibration isn't too sensitive to that, provided that $\mu > 0.1$. So, the speed of mean-reversion coming from inertia is $\Phi = \frac{\phi}{\mu} = 2/\text{year}$, and the overshooting of flows on impact is, using $\kappa = 1.5$ year for illustration:

$$b = \frac{1-\mu}{\mu\zeta + \phi\kappa} = \frac{1-0.5}{0.5 \cdot 0.2 + 1 \cdot 1.5} \simeq 0.3.$$

The immediate price impact is $\frac{1}{\zeta} + b = 5.3$, while the permanent price impact is $\frac{1}{\zeta} = 5$. So, the temporary bump $b \ll \frac{1}{\zeta}$ is pretty negligible in the big picture. As the price decays as $bf_t (1-\Phi)^t$, the premium is $b\Phi f = 0.3 \cdot 2 \cdot 0.5\% = 0.3\%$ (if $f_t = 0.5\%$), again a small premium. A higher inertia (lower ϕ) creates a bigger difference between long run and short term price impact. So, examining inertia across investor classes is a useful avenue for future research.

Impact of anticipated and unanticipated flows when funds are partially inert We next generalize the price as present value formula (16) and the price impact with inertia (166).

Proposition 18. (Price impact with inertial funds) *When funds exhibit inertia, the price impact of inflows df_s is:*

$$p_t = \frac{f_{-\infty}}{\zeta} + \sum_{s=-\infty}^{\infty} G(t-s)\, \mathbb{E}_t [\Delta f_s], \tag{167}$$

where

$$G(\tau) = \begin{cases} \left(\frac{1}{\zeta} + b\right) \frac{1}{(1+\rho)^\tau} & \text{if } \tau < 0, \\ \frac{1}{\zeta} + b(1-\Phi)^\tau & \text{if } \tau \geq 0. \end{cases} \tag{168}$$

and b is in Proposition 17. When there is no inertia, $b = 0$.

Proof. This can be checked by the "plug and verify" method. □

Heterogeneity in inertia across funds. One can generalize this model to the case of heterogeneous inertial funds. Things are particularly tractable when ϕ_i is the same across funds i (but μ_i, ζ_i, κ_i could be different): then (166) holds, with more complex expressions for b and Φ.

E.13 Calibration of the general equilibrium model of Section 5: Details

Here we provide a detailed justification of the calibration in Section 5.5.

E.13.1 From discrete to continuous time

Denote with bolded symbols the continuous-time version of the parameters, and Δt the calendar value of a time interval. Then, as they have units of $[\text{Time}]^{-1}$, we have

$$\phi = \boldsymbol{\phi}\Delta t, \qquad \delta = \boldsymbol{\delta}\Delta t, \qquad \rho = \boldsymbol{\rho}\Delta t$$

[108] We use a "continuous time" calibration: The expressions work in continuous time (which makes calibration easier, replacing expressions like $(1-\phi)^t$ by $e^{-\phi t}$.

but as $\boldsymbol{\kappa}$ has unit of [Time], we have

$$\kappa = \boldsymbol{\kappa} (\Delta t)^{-1} \tag{169}$$

As it is unitless, ζ is the same in discrete and continuous time,

$$\zeta = \boldsymbol{\zeta}$$

.

More subtilely, $\boldsymbol{\sigma}_f$ has the units of $[\text{Time}]^{-1/2}$ so

$$\sigma_f = \boldsymbol{\sigma}_f (\Delta t)^{1/2}$$

and similarly for $\boldsymbol{\sigma}_d$.

E.13.2 Calibration steps

The calibration steps are the following. We express things in annualized values, but we use the correspondence in Section E.13.1 to go between discrete vs continuous time notions.

1. For Tables 4 and 5, first we set $\gamma = 2, g = 2\%, \sigma_y = 0.8\%, \sigma_D = 5\%, r_f = 1\%, \sigma_f = 2.8\%$ (in this appendix we use σ_f for the volatility of the innovations to $\tilde{f}_t$). Next we impute β from r_f given γ, g, σ_y^2. Then we calculate $\sigma_r^2, \bar{\pi}, \delta$ and set $\phi_f = 4\%, \zeta^M = 0.2, \kappa = 1, \phi_f = 9\%, \theta = 0.85$, which jointly imply $\zeta = \zeta^M - \kappa\phi_f = 0.16, m_p = \frac{\zeta - \kappa\delta}{1-\theta} = 0.84$. Last we calculate $\rho, b_f^p, b_f^\pi, \sigma_b$.

2. In Table 5, we use $\hat{r}_t \coloneqq r_t - \mathbb{E}_{t-1}[r_t] = \varepsilon_t^D + b_f^p \varepsilon_t^f$: so, the share of the variance of excess stock returns that is due to flows is $\frac{cov(\hat{r}, b_f^p \varepsilon_t^f)}{var(\hat{r})}$. Likewise, the share of the variance of stock returns due to fundamentals is $\frac{cov(\hat{r}, \varepsilon_t^D)}{var(\hat{r})}$. The two shares add up to 1.

3. For Table 6a, the model implied mean of P/D is

$$\mathbb{E}\left[\frac{P}{D}\right] = \mathbb{E}\left[\frac{\mathrm{e}^{p_t}}{\delta}\right] = \frac{1}{\delta}\mathbb{E}\left[\exp\left(b_f^p \tilde{f}_t\right)\right] = \frac{1}{\delta}\exp\left(\frac{1}{2}\left(b_f^p\right)^2 \frac{\sigma_f^2}{1-(1-\phi_f)^2}\right), \tag{170}$$

   the log mean of P/D is

$$\exp\left(\mathbb{E}\log P/D\right) = \frac{1}{\delta}, \tag{171}$$

   and the variance of log D/P is

$$\text{Var}\left(\log D/P\right) = \text{Var}\left(p_t\right) = \frac{\left(b_f^p\right)^2 \sigma_f^2}{1-(1-\phi_f)^2}. \tag{172}$$

   Up to now, everything is in units of continuous time.

4. For Table 6b, we simulate the model with $\Delta t = \frac{1}{12}$ over 72 years with $N = 1000$ simulations. We report the mean and 95% confidence interval for the slopes, the mean 8-lag Newey-West standard errors and the mean R^2 across the $N = 1000$ simulations as the model-generated results.

(a) We generate $T = \frac{72}{\Delta t}$ periods (72 years) of i.i.d. innovations $\varepsilon_t^D \sim \mathcal{N}\left(0, \sigma_D^2 \cdot \Delta t\right), \varepsilon_t^f \sim \mathcal{N}\left(0, \sigma_f^2 \cdot \Delta t\right)$ with $\Delta t = \frac{1}{12}$ and calculate the path of $d_t, \tilde{f}_t$ from

$$d_t - d_{t-1} = g \cdot \Delta t + \varepsilon_t^D - \frac{1}{2}\sigma_D^2 \cdot \Delta t, \tag{173}$$

$$\tilde{f}_t - \tilde{f}_{t-1} = -\phi_f \Delta t \cdot \tilde{f}_{t-1} + \varepsilon_t^f, \tag{174}$$

for a total of $N = 1000$ times.

(b) We calculate the time series for the log dividend, the deviation of prices from their rational average, the log price, the log dividend price ratio, and the equity premium as follows:

$$\log D_t = d_t, \tag{175}$$

$$p_t = b_f^p \tilde{f}_t, \tag{176}$$

$$\log P_t = \log D_t + p_t - \log \delta, \tag{177}$$

$$\log D_t/P_t = \log D_t - \log P_t, \tag{178}$$

$$\pi_t = \bar{\pi} + b_f^\pi \tilde{f}_t. \tag{179}$$

(c) For the predictive return regressions, we collapse the data to a yearly frequency by taking the price and the dividend paid in the last month of the year. Then we calculate the 1-year, 4-year and 8-year cumulative returns, and run 1-year, 4-year and 8-year predictive regressions on the collapsed yearly data. (For the 4-year and 8-year horizons, the regressions have overlapping windows when calculating returns.) We report the 8-lag Newey-West standard errors for all three regressions.

i. In both the data and the simulations, returns are calculated assuming that the investor pockets the monthly dividends without reinvesting them at the risk-free rate. For example, the return at the 1-year horizon is the month-12 post-dividend P_{12} plus an average of D_1 to D_{12} (approximating $\int_0^1 D_\tau d\tau$ in continuous time) divided by the month-0 post-dividend price P_0.

ii. We generate the *cumulative* return for the 4-year and 8-year horizons by compounding 1-year return $(r_{t,t+4} = (1 + r_{t,t+1})(1 + r_{t+1,t+2})(1 + r_{t+2,t+3})(1 + r_{t+3,t+4}) - 1)$ for the predictive regressions.

References for Online Appendix

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, "Hedge fund stock trading in the financial crisis of 2007–2009," *The Review of Financial Studies*, 2012, *25* (1), 1–54.

Benzaquen, Michael, Iacopo Mastromatteo, Zoltan Eisler, and Jean-Philippe Bouchaud, "Dissecting cross-impact on stock markets: An empirical analysis," *Journal of Statistical Mechanics: Theory and Experiment*, 2017, *2017* (2), 023406.

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

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts, "On the importance of measuring payout yield: Implications for empirical asset pricing," *The Journal of Finance*, 2007, *62* (2), 877–915.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, "Regression Discontinuity and the Price Effects of Stock Market Indexing," *Review of Financial Studies*, 2014, *28* (1), 212–246.

Chinco, Alex, Samuel Hartzmark, and Abigail Sussman, "Necessary Evidence For A Risk Factor's Relevance," 2020. Working Paper.

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

Duffie, Darrell, "Presidential Address: Asset Price Dynamics with Slow-Moving Capital," *Journal of Finance*, 2010, *65* (4), 1237–1267.

Frazzini, Andrea, Ronen Israel, and Tobias J Moskowitz, "Trading costs," *Available at SSRN 3229719*, 2018.

Gabaix, Xavier and Ralph SJ Koijen, "Granular instrumental variables," *Available at SSRN 3368612*, 2020.

Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and H Eugene Stanley, "A theory of power-law distributions in financial market fluctuations," *Nature*, 2003, *423* (6937), 267–270.

Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and H Eugene Stanley, "Institutional investors and stock market volatility," *The Quarterly Journal of Economics*, 2006, *121* (2), 461–504.

Greenwood, Robin and Andrei Shleifer, "Expectations of returns and expected returns," *The Review of Financial Studies*, 2014, *27* (3), 714–746.

Hasbrouck, Joel, *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*, Oxford University Press, 2007.

Koijen, Ralph SJ and Motohiro Yogo, "A demand system approach to asset pricing," *Journal of Political Economy*, 2019, *127* (4), 1475–1515.

Koijen, Ralph SJ, Robert J Richmond, and Motohiro Yogo, "Which investors matter for global equity valuations and expected returns?," *Working Paper*, 2019.

Kyle, Albert S., "Continuous Auctions and Insider Trading," *Econometrica*, 1985, *53* (6), 1315–1335.

Lillo, Fabrizio, Szabolcs Mike, and J. Doyne Farmer, "Theory for long memory in supply and demand," *Phys. Rev. E*, Jun 2005, *71*, 066122.

Lucas, Robert E. Jr., "Asset Prices in an Exchange Economy," *Econometrica*, 1978, *46* (6), 1429–1445.

Madhavan, Ananth, Matthew Richardson, and Mark Roomans, "Why do security prices change? A transaction-level analysis of NYSE stocks," *The Review of Financial Studies*, 1997, *10* (4), 1035–1064.

Shleifer, Andrei, "Do Demand Curves for Stocks Slope Down?," *Journal of Finance*, 1986, *41* (3), 579–590.

Torre, N and Mark J Ferrari, "The market impact model," *The Barra Newsletter*, 1998, *165.*

Wurgler, Jeffrey and Ekaterina Zhuravskaya, "Does Arbitrage Flatten Demand Curves for Stocks?," *Journal of Business*, 2002, *75* (4), 583–608.