

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

Xavier Gabaix and Ralph S.J. Koijen*

March 3, 2022

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 in response to changing market conditions. As a result, the price elasticity of demand of the aggregate stock market is small, and 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 develop a new measure of capital flows into the market, consistent with our theory. We relate it to prices, macroeconomic variables, and survey expectations of returns.

We 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 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, Manav Chaudhary, Aditya Chaudhry, Antonio Coppola, Zhiyu Fu, 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, Francesco Franzoni, Robin Greenwood, Arpit Gupta, Valentin Haddad, Lars Hansen, Sam Hanson, John Heaton, Tim Johnson, Arvind Krishnamurthy, Spencer Kwon, John Leahy, Daniel Lewis, Hanno Lustig, Alan Moreira, Knut Mork, Toby Moskowitz, Stefan Nagel, Jonathan Parker, Lasse Pedersen, Joel Peress, Monika Piazzesi, Jean-Charles Rochet, Ivan Shaliastovich, Andrei Shleifer, Jeremy Stein, Johannes Stroebl, 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 suggesting that this is because of flows and demand shocks 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 (our estimates are between \$3 and \$8, and we will use \$5 for simplicity 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 and demand shocks affect 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 this 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. Moreover, the model aggregates well. If different investors have different elasticities, the total market elasticity is the size-weighted elasticity of market participants. Importantly, the correct measure of size is the share of equity they hold. The model also clarifies how to measure net flows into the aggregate stock market (even though for every buyer there is a seller), which guides the empirical analysis. Moreover, it extends readily to an infinite horizon: in that case, 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.

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”

¹The price impact is linear and symmetric: 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 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 necessarily 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. We also construct a measure of capital flows into the market. We find that this measure is strongly correlated with realized returns and survey expectations of returns, but it is only weakly correlated with macroeconomic growth.

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 empirically. Relatedly, it is hard to find investors who could act as macro arbitrageurs. For instance, hedge funds are relatively small (they hold less than 5% of the equity market), and they tend to reduce their equity allocations in bad times (due to outflows and binding 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 implies that the demand elasticity of most investors is quite small or that investors experience nearly identical demand shocks (as if they were to disagree, we would see large flows in elastic markets), something which may be implausible. Third, a large literature estimates demand elasticities for individual stocks using a variety of methodologies, where the latest estimates of this “micro” demand elasticity are approximately 1 (we provide complete references 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 1. Consistent with this reasoning, a new literature explores elasticities for “factors” in the US, such as size and value, and finds elasticities of around 0.2. Hence, in light of this existing evidence, our low macro elasticity may be less surprising.

Suppose that the “inelastic markets hypothesis” is true; why do we care? First, investor-specific flows and demand shocks are quantitatively impactful. 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. This suggests a research program in which determinants of asset prices can be traced back to measurable demand shocks and flows of 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 quantify 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 around \$2.² 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 illiterate. But 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 therefore of the volatility and increase in the value 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 demand shocks and 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 the price-dividend ratio at different horizons. We also show how the model can be used to match the strong correlation between prices and subjective beliefs about long-term growth (Bordalo et al. (2020)), even if fluctuations in beliefs have only a modest impact on actions (Giglio et al. (2021a)), as the resulting flows are amplified in inelastic markets. 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 it relies on an observable force, flows in and out of equities.

We also show how to connect 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

²The estimate is tentative, in part as it relies on estimates of the rationality of the consumer after the buybacks.

correlate sensibly with other variables, such as prices 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 questions such as “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).

Literature review Our paper is about the macro elasticity, in contrast to the micro elasticity estimated in the literature, including Shleifer (1986), Harris and Gurel (1986a), Wurgler and Zhuravskaya (2002), and Duffie (2010).³ We summarize the evidence on existing elasticity estimates in more detail in Section 2.4.

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. They 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, for example Warther (1995).⁴ In addition, an active literature studies the impact of mutual fund and ETF flows on the cross-section of equity prices, for instance Frazzini and Lamont (2008), Lou (2012), Ben-David et al. (2018), Dou et al. (2020), and Dong et al. (2021). 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 to think about causal inference at the level of the aggregate stock market via GIV. 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 horizons that they consider.⁵

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), who emphasize flows coming from the supply of shares by firms), 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 US equity 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

³A growing literature studies elasticities in global financial markets, see for instance Dierker et al. (2016) and Charoenwong et al. (2020).

⁴See also Edelen and Warner (2001), Goetzmann and Massa (2003), and Ben-Rephael et al. (2012).

⁵Deuskar 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, their results do show that flows explain a large fraction of market fluctuations, which is consistent with the inelastic markets hypothesis.

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, it avoids CARA restrictions on investor preferences) 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 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), Koijen and Yogo (2020) estimate a global demand system across global equity and bond markets to understand exchange rates, bond prices, and equity prices across countries. 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 on equity shares and potential macro arbitrageurs such as broker dealers and hedge funds. It also summarizes the existing literature on elasticity estimates. 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 contains 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 specializes the basic model of Section 3 as it endogenizes the interest rate and links cash flows to production and consumption. Section 6 discusses how the effectiveness of government policy and corporate finance change with inelastic markets. 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 $X_E := \frac{1}{N} \sum_{i=1}^N X_i$, $X_S := \sum_{i=1}^N S_i X_i$, $X_\Gamma := X_S - X_E$ 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: $\mathbb{E}_\omega [X_i] := \frac{\sum_i \omega_i X_i}{\sum_i \omega_i}$.

2 Data and Suggestive Facts on Equity Shares and Flows

In this section we document several stylized facts and discuss how they are related to our model and to traditional, elastic asset pricing models. 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.

After discussing the data construction in Section 2.1, we document that institutions often have quite stable equity shares in Section 2.2, and relatedly we seek to identify investors with elastic demand for the aggregate stock market in Section 2.3. 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.

2.1 Data sources and construction

We summarize the data sources that we use and define some of the key variables. We leave a detailed description for Appendix C.

We use sector-level data from the Flow of Funds (FoF) on holdings of equities and bonds as well as flows into both asset classes. Flows are differences in levels adjusted for mechanical valuation effects. We compute total bond holdings as the sum of Treasury and corporate bond holdings, and analogously for flows. As the FoF reports combined values of holdings and flows of foreign and US assets (except for Treasuries), we adjust these series (Appendix C.1.3). We assume that the flows transact at end-of-period prices. The sample is quarterly from 1993 to 2018 and we use the June 2019 vintage of the FoF data.⁶

We use monthly disaggregated data on assets under management, the share invested in US equities, and flows from Morningstar for mutual funds and ETFs that are domiciled in the US and that have the US dollar as the base currency. We select the funds in Morningstar’s US category groups “US Equity,” “Sector Equity,” “Allocation,” and “International Equity.”⁷ We use the sample from 1993 to 2019 for mutual funds and from 2002 to 2019 for ETFs.⁸

For state and local pension funds, we use data from the Center for Retirement Research at Boston College. The sample is from 2002 to 2019. We use data on the share invested in equities and fixed income as well as target holdings in equities and fixed income (including cash). State and local pension funds report once a year (although in different quarters). We use a fund’s actual and target allocation to equity and fixed income and scale it so that the sum of the shares equals 100% for each fund.

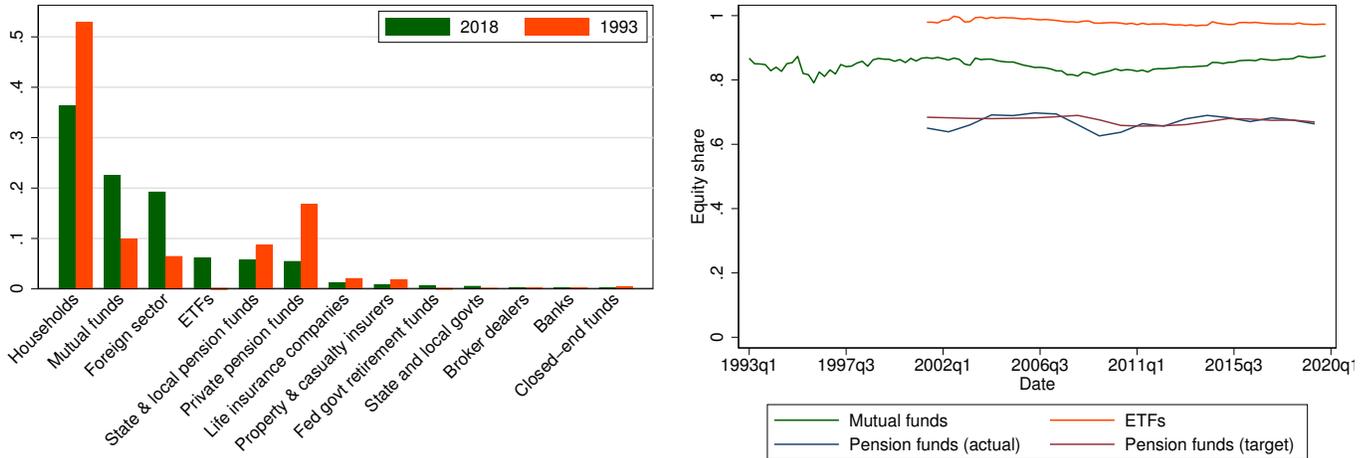
We use disaggregated data on equity holdings by institutional investors via form 13F filings. We source the 13F filings from FactSet and the construction is as in Koijen et al. (2019). The sample

⁶Data of different vintages can be downloaded from this website.

⁷We remove fund of funds in our analysis to avoid double counting.

⁸We omit a small number of fund-quarters in which the US equity share exceeds 300% or is lower than -300%, as these may be data errors.

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 using Flow of Funds data. 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.



is from 1999 to 2019.

We use quarterly data on real GDP growth from the St. Louis Federal Reserve Bank FRED database, series GDPC1. Data on returns with and without dividends are from the Center for Research in Security Prices. We use the monthly, value-weighted return with and without dividends to compute the monthly dividend payment.

Lastly, 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, and the resulting series has some gaps.

2.2 Institutions often have a quite stable 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) based on FoF data. 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. We aggregate different investors in a given sector using the relative sizes of their equity portfolios as opposed to assets under management, consistently with our theory (see the discussion around 15). To appreciate the importance of this difference, consider an economy with only pure equity and pure bond funds that have the same amount of assets under management. The equity-weighted equity share equals 100% while the asset-weighted equity share equals only

50%. As the relative size of equity and bond assets move, so will the asset-weighted equity share. Yet, the equity-weighted share will be a constant 100%. It is the equity-weighted equity share that is relevant per our theory.

The plot shows that equity shares are quite stable over time for broad classes of investors. This is consistent with many institutions having a rather rigid mandate to maintain a stable equity share. In the model that we introduce in Section 3, this mandate rigidity will be captured by a low elasticity (κ) of funds' asset location to the expected return on equities. In recent work, Cole et al. (2021) show that a large fraction of households⁹ also have a high average equity share at 79.2% with little variation over time (the equity-weighted equity share only drops to 76.4% at the end of 2008). This stability is in part explained by the introduction of target date funds.

2.3 In search of macro arbitrageurs

Figure 1 shows that the equity shares of large groups of investors, such as mutual funds, ETFs, and pension funds, are stable over time. As the foreign sector consists of similar institutions, this fact naturally raises the question of who carries out arbitrage across asset classes or, equivalently, which group of investors aggressively times the market. In the survey that we discuss in the introduction, two investor sectors are frequently mentioned: hedge funds and broker dealers.¹⁰

As Figure 1 shows, broker dealers are very small and hold less than 0.5% of the equity market directly. So while perhaps important for the micro elasticity, broker dealers are not well-positioned to absorb large equity flows over longer periods of time. The hedge fund sector is also quite small, with holdings below 4% of the equity market in long positions going into the financial crisis. Ben-David et al. (2012) document two important facts. First, hedge funds *sold* a large fraction of their equity holdings during the financial crisis, averaging to 3.06% per quarter from 2007.Q3 to 2009.Q1. Given their small size, this corresponds to selling on average 0.1% of the market each quarter (or 0.7% in total). Redemptions and leverage constraints explain about 80% of this decline in equity holdings. Second, flows across sectors are small. Ben-David et al. (2012) decompose the market into hedge funds, mutual funds, short sellers, other institutional investors (e.g., pension funds and insurance companies), and non-institutional investors (e.g., households). Measured as a fraction of the market, these investor sectors sell or buy on average just 0.25% of the market per quarter. We extend these calculations using data from the FoF for the technology crash in 2000-2002 and the 2008 global financial crisis in Appendix D.3. As a fraction of the market, flows between groups average to at most 0.5% of the market.

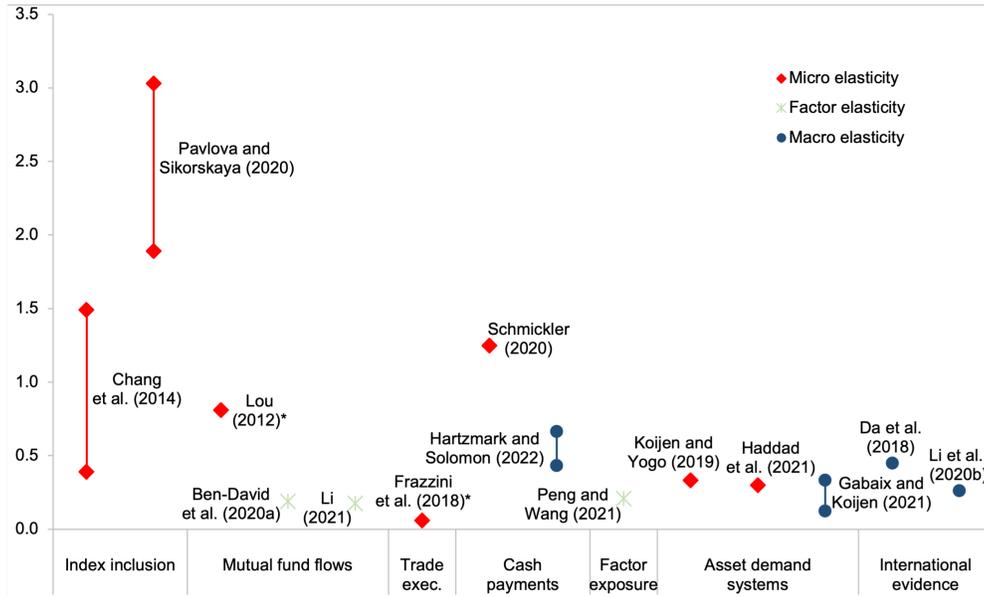
In summary, many funds appear to have fairly tight mandates, hedge funds do not appear to arbitrage the aggregate stock market and amplify demand shocks during severe downturns, and flows between sectors are small.

The small flows across sectors has implications for the properties of demand shocks, which are shocks to investors' beliefs or risk appetite, given the elasticity of demand. The signature of elastic demand is that disagreement among investors is associated with large flows and quantity movements. As flows are small, theories featuring elastic demand imply that investors should agree almost perfectly in their beliefs about expected growth rates and their riskiness, and also have similar risk aversion. In inelastic markets, in contrast, there can still be large common shocks to beliefs, for

⁹Their sample appears to be representative of the middle 80% of the retirement wealth distribution of retirement investors between age 25 and 65.

¹⁰While 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.

Figure 2: Elasticity estimates in the existing literature. The figure reports elasticity estimates in the existing literature for individual stocks, factors such as size and value, and the aggregate stock market. The elasticity is defined as the percent change in demand per percent change in prices. We discuss in footnote 12 and Appendix G.5 how to interpret the trade-level estimates of Frazzini et al. (2018); here, we simply report the “prima facie” estimates. In the bottom of the figure, we summarize the methodology used to identify the elasticities. For papers with a *, we compute the elasticity based on the tables reported in the paper ourselves.



instance as during the 2008 financial crisis, but there is much more scope for disagreement.¹¹ This second interpretation of financial markets may be more consistent with the data on beliefs, which points to significant fluctuations in disagreement over time (Giglio et al. (2021a)).

2.4 The micro and macro elasticity of markets: Summary of existing evidence

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 market (which describes how much the relative price of two stocks changes if one buys \$1 of one, and sells \$1 of the other). A more recent literature studies multipliers at different levels of aggregation.

We summarize this literature in Figure 2 and Table 1. In the figure, we report recent demand elasticity estimates, which are defined as the percent change in aggregate demand per percent

¹¹To make this more concrete, using the notation of the next section, consider the simple decomposition of demand $\Delta q_{it} = -\zeta \Delta p_t + f'_{it}$. If markets are as elastic as in standard models, say $\zeta = 10$, then $f'_{it} = \Delta q_{it} + 10\Delta p_t$. As the volatility of Δq_{it} is modest, demand shocks are largely dominated by the second term, $10\Delta p_t$, and almost perfectly correlated. This leaves little room for disagreement among investors, even though this is widely documented in beliefs data, as the signature prediction of a model with elastic markets and belief disagreement is the presence of large flows coupled with small price changes. When markets are inelastic, say $\zeta = 0.2$, then $f'_{it} = \Delta q_{it} + 0.2\Delta p_t$. Demand shocks still contain a large common component, but the correlation between demand shocks is much lower and there is more scope for disagreement.

Table 1: Multiplier estimates in the existing literature. The table reports multiplier estimates in the existing literature for individual stocks (Panel A), factors such as size and value (Panel B), and the aggregate stock market (Panel C). The multiplier is defined as the percent change in prices per percent change in shares outstanding purchased or sold by an investor. [†]We discuss in footnote 12 and Appendix G.5 how to interpret the trade-level estimates of Frazzini et al. (2018) and Bouchaud et al. (2018); here, we simply report the “prima facie” estimates.

Panel A: Micro multiplier		
	Methodology	Multiplier
Chang, Hong and Liskovich (2014)	Index inclusion	0.7 to 2.5
Pavlova and Sikorskaya (2020)	Index inclusion	0.3 to 0.5
Schmickler (2020)	Dividend payouts	0.8
Lou (2012)	Mutual fund flows	1.2
Frazzini et al. (2018), Bouchaud et al. (2018)	Trade-level permanent price impact	15 [†]
Panel B: Factor-level multiplier		
Ben-David, Li, Rossi and Song (2020a)	Morningstar ratings change	5.3
Peng and Wang (2021)	Fund flows	4.8
Li (2021)	Fund flows+SVAR	5.7
Panel C: Macro multiplier		
Hartzmark and Solomon (2022)	Dividend payouts, US	1.5-2.3
Da, Larrain, Sialm and Tessada (2018)	Pension fund rebalancing, Chile	2.2
Li, Pearson and Zhang (2020b)	IPO restrictions, China	2.6-6.5

change in prices. In the bottom of the figure, we summarize the methodology used to identify the elasticities.

In Panel A of Table 1, we provide a summary of recent estimates of the micro multiplier, which is the percent change in prices when an investors purchases a certain fraction of the shares outstanding in a particular company, while controlling for movements in the aggregate market.¹² While there is a range of estimates, the order of magnitude of the multiplier is around 1. That is, buying 1% of the shares outstanding of a given stock results makes its price increase by around 1%.

¹²Also, 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 (Bouchaud et al. (2018); Frazzini et al. (2018)), which may make our estimate of 5 seem moderate. Microstructure results are typically couched in a form such as “buying 2.5% of the daily volume of a stock creates a permanent price increase of 0.15%”. At first glance, values in this range might appear to imply a small price impact. However, they work out to a large price impact multiplier of $M = 15$: with 250 days of trading in a year, and a 100% per year turnover, the trade in our example would represent a purchase of $\frac{2.5\%}{250} = 0.01\%$ of the market capitalization of a stock, so that the impact of 0.15% on the price results in a multiplier of 15. The interpretation of this kind of microstructure estimates requires some caution, as we discuss in Section G.5. To sum up, a microstructure estimate of 15 may have the following interpretation: in inelastic markets with a micro elasticity equal to 1, 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). These microstructure estimates are also themselves to be taken with caution, since identification tends to be difficult as trades are not exogenous to prices. Using high frequency data with a GIV-based identification may be a promising way to enrich identification procedures in microstructure.

In addition, several recent studies (sometimes motivated by the present paper) have looked at the “factor-level” multiplier, which is the price impact if an investor buys a fraction of the shares outstanding of a cross-sectional factor such as size or value. We report those estimates in Panel B. The studies report a multiplier that is substantially above 1 and closer to 5. In Panel C, we report recent estimates of the “macro multiplier,” the parameter of interest in this paper, for various stock markets.¹³ Once again, the multiplier estimates are well above 1. Equivalently, the macro elasticity, which is the inverse of the multiplier, is well below 1.

In addition to these estimates, Haddad et al. (2021) find that micro multipliers increase in firm size and, relatedly, Li and Lin (2022a) find that multipliers increase in the level of aggregation. They find a micro multiplier around 2 and a multiplier around 5 at the factor level.

Another important question is whether prices revert back to their levels prior to the shock and how long this takes. While this question is explored only in a subset of the papers, Lou (2012) finds that it takes between six to nine quarters for prices to revert back. Ben-David et al. (2020a) find that it takes about two to three years for prices to revert back at the factor level. Pavlova and Sikorskaya (2020) show how stocks included or excluded in an index have differences in average returns from one year up to five years. Taken together, the existing evidence in the literature suggests a micro multiplier around 1 (so, a micro elasticity around 1), and a factor or macro multiplier that is well above 1 (so, a macro elasticity below 1). In addition, the estimated effects are not confined to high frequency effects, and are of relevance to macro finance theories.

How do these estimates compare to the elasticities implied by standard asset pricing models? It is well known (e.g. Petajisto (2009)) that the micro elasticity in standard models is very large, of the order of 5000 or above. This implies that the micro multiplier (the inverse of the micro elasticity) is essentially zero and “demand curves are virtually flat.” Based on the estimates reported in Table 1, the models are several orders of magnitudes off in terms of the micro elasticity.

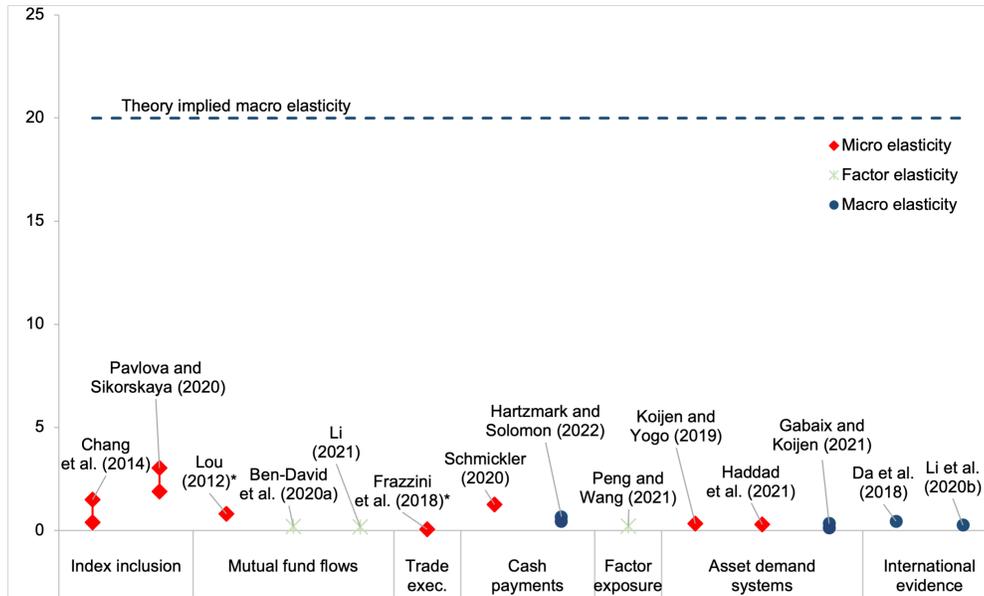
Our focus is on the macro elasticity and we compute it for various frictionless neoclassical asset pricing models in Section F.4.¹⁴ The summary is that in traditional, elastic asset pricing models the macro elasticity is around 10 to 20, leading to a multiplier around 0.1 to 0.05. As any two stocks are closer substitutes than stocks and bonds, the micro multiplier is much lower than the macro multiplier in standard asset pricing models. However, the micro multiplier as estimated in the literature (see Panel A) is already an order of magnitude larger than the macro multiplier implied by standard asset pricing models. The macro multiplier estimates are even larger, which deepens the disconnect between existing estimates and asset pricing models. In Figure 3, we provide a direct comparison of the empirical estimates to both the macro elasticities implied by theoretical models. The micro elasticity implied by theory is two orders of magnitude larger.

The profession’s view on the macro elasticity and the underlying mechanism While the disconnect between the empirical estimates and asset pricing models follows from the existing literature, these facts have typically not been targeted in macro-finance asset pricing models. In fact, as we will discuss now, this evidence does not appear to be widely known or accepted in the profession.

¹³In interpreting the multipliers in the context of various theories, it is important to note that some shocks are in principle anticipated (for instance, those related to dividend payouts), while others are unanticipated (for instance, those related to mutual fund flows). In traditional models, and also in our theory, anticipated flows have a smaller impact on prices compared to unanticipated flows.

¹⁴We discuss the elasticity in the models of Lucas (1978), Bansal and Yaron (2004), Barro (2006), Gabaix (2012), and the link between our findings and Johnson (2006).

Figure 3: Comparison of elasticity estimates to the elasticities implied by frictionless neoclassical theories. We compare the empirical elasticity estimates reported in Figure 2 to the elasticity implied by theory, which is around 20 (depending on the model). The theoretical micro elasticity is two orders of magnitudes larger and omitted for clarity.



We quantify this via two surveys. We provide a detailed discussion in Section E and summarize the main insights here. 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 macro multiplier, which we estimate to be around $M = 5$. In both surveys, the median answer was $M = 0$: surveyed economists, logically enough, rely on the traditional asset pricing model in which prices are unperturbed by flows. The median positive answer was $M = 0.01$.¹⁵ Hence, surveyed economists’ views are in line with the traditional model, but far from the estimates reported in the empirical literature, and the new estimates we provide.

We also asked about the sector supposedly providing elasticity to the market to be able to explore the mechanism. The two most common responses were hedge funds and broker dealers. As discussed before, those sectors are unlikely to provide elasticity to the aggregate market, in particular during times of stress.

¹⁵The 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 E provides further details.

3 The Inelastic Markets Hypothesis: Theory

We now provide a model that we think is more realistic to think concretely about the determinants of stock demand, and about how flows impact prices. 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.

3.1 Two-period model

There is a representative stock in fixed supply of Q shares, with an endogenous price P . The economy lasts for two periods $t = 0, 1$. The dividend D is paid at time 1. We call $\pi = \frac{D^e}{P} - 1 - r_f$ the equity premium (with $D^e := \mathbb{E}[D]$ the expected dividend at time 0 and r_f the risk-free rate), $\bar{\pi}$ the average equity premium, and $\hat{\pi} := \pi - \bar{\pi}$ the deviation of the equity premium from its average. There is also a riskless bond with time-0 price equal to 1 (we endogenize the risk-free rate in Section 5).¹⁶

A representative consumer invests into stocks and bonds via I institutions or funds.¹⁷ We call W_i fund i 's wealth (or equivalently assets under management) and Q_i the number of stock market shares it holds. Therefore the fraction of fund i 's wealth invested in equities is $\frac{PQ_i}{W_i}$. We assume that fund i 's demand for stocks is given by a mandate, saying that it should have a fraction invested in equities equal to:¹⁸

$$\frac{PQ_i}{W_i} = \theta_i e^{\kappa_i \hat{\pi}}, \quad (1)$$

while the rest is in the riskless bond. In the simplest case, $\kappa_i = 0$, fund i has a fixed mandate to invest a fraction $\theta_i \geq 0$ of its wealth in equities. When $\kappa_i > 0$ the fund allocates more in equities when they have higher expected excess returns (hence, κ_i 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.^{19/20/21} We use the index $i = 0$ for a special fund, a “pure bond fund” that only holds bonds (so, its θ_i and κ_i are 0).

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

¹⁶Here, flows move equity prices but not bond prices. In the general equilibrium version of Section 5, this happens because the consumer's demand is infinitely elastic with respect to bond prices. We sketch the case where both equity and bond demands are inelastic in Section G.1: the economics is similar, replacing the elasticity by a matrix of own- and cross-elasticities.

¹⁷Those funds act competitively, i.e. are price takers.

¹⁸We write the mandate in “number of shares,” but it is equivalent to a “fraction of assets invested in equity” formulation.

¹⁹This 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 as it depends on how households manage their own portfolios. Parker et al. (2020) argue that the growth of target date funds made the market more elastic. In the our notation, target date funds have $\kappa_i = 0$.

²⁰Buffa et al. (2019) explore the implications of tracking error constraints on asset prices.

²¹The mandate does not feature volatility, as volatility is not crucial here to obtain demand curves (though volatility is crucial for that in the traditional model). One could easily write extensions where the allocation decreases in volatility. In the dynamic model, we add a demand shock that can include volatility terms.

have an impact.

The elasticity of demand for stocks of a fund We use bars to denote values at time $t = 0^-$, before any shocks. At that time 0^- , fund i has wealth \bar{W}_i , and holds \bar{Q}_i 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.

At time 0, the representative household invests ΔF_i extra dollars in each fund i (taking those dollars from the pure bond fund), which represents a fractional inflow $f_i = \frac{\Delta F_i}{\bar{W}_i}$. An outflow corresponds to $\Delta F_i < 0$. We study the impact of this on the aggregate market, independently of the reasons for the flows, which may be rational or behavioral. We also assume that there may be a change d in the value of expected fundamentals. We call q_i and d the fractional deviations of equity demand and of the expected dividend from their baseline values:

$$q_i = \frac{Q_i}{\bar{Q}_i} - 1, \quad d = \frac{D^e}{\bar{D}^e} - 1. \quad (2)$$

The next proposition gives the change in demand by fund i . Its proof is in Appendix A. We perform the analysis for small disturbances f_i, d , and hence small p, q_i , here and throughout the paper.²²

Proposition 1. (Demand for aggregate equities in the two-period model) *Fund i 's demand change (compared to the baseline) is, linearizing:*

$$q_i = -\zeta_i p + \kappa_i \delta d + f_i, \quad (3)$$

where δ is the baseline dividend-price ratio, and ζ_i is the elasticity of equity demand by fund i ,

$$\zeta_i = 1 - \theta_i + \kappa_i \delta. \quad (4)$$

The aggregate elasticity of demand for stocks, and the “representative mixed fund” We now move from fund-level demand to the aggregate demand for stocks, which is $Q = \sum_i \bar{Q}_i (1 + q_i)$. We call $W_i^{\mathcal{E}}$ the equity holdings (in dollars) of fund i , and S_i its share of total equity holdings:

$$W_i^{\mathcal{E}} = Q_i P = \theta_i W_i, \quad S_i = \frac{\bar{W}_i^{\mathcal{E}}}{\sum_j \bar{W}_j^{\mathcal{E}}} = \frac{\bar{Q}_i}{\sum_j \bar{Q}_j}. \quad (5)$$

Finally, for a given variable x_i (with $i = 1 \dots I$), we define x_S to be its equity-holdings weighted mean:

$$x_S := \sum_i S_i x_i. \quad (6)$$

So, the aggregate demand change is:

$$q = \frac{\Delta Q}{Q} = \frac{\sum_i \bar{Q}_i q_i}{Q} = \sum_i S_i q_i = q_S.$$

²²Following common practice in macro-finance, we do Taylor expansions of the leading terms, omitting the formal mentions of $O(\cdot)$ terms.

To derive an expression for it, we take the individual demand curves (3), and consider their equity-holdings weighted average, which gives the (linearized) aggregate demand curve for equities:

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

Proposition 2 sums this up.

Proposition 2. (Aggregate demand for aggregate equities in the two-period model) *The aggregate demand for equities is*

$$q = -\zeta p + \kappa \delta d + f, \quad (7)$$

where $\zeta = \zeta_S = \sum_i S_i \zeta_i$ is the equity-holdings weighted demand elasticity of all funds i , and likewise for the other quantities:

$$\theta = \theta_S, \quad \kappa = \kappa_S, \quad \zeta = \zeta_S, \quad f = f_S. \quad (8)$$

In particular, ζ is the macro elasticity of demand:

$$\zeta = 1 - \theta + \kappa \delta. \quad (9)$$

Hence, the universe of equity-holding funds in the model aggregates (up to second order terms in f_i and d) to a “representative mixed fund” with wealth $W = \sum_{i=1}^I W_i$, and whose mandate is to hold an equity share $\frac{PQ}{W} = \theta e^{\kappa \hat{\pi}}$.

The “aggregate flow into equities” is non-zero even though “for every buyer there is a seller” The equity-share weighted flow $f_S = \sum_i S_i f_i$ in (8) can also be expressed as²³

$$f_S = \frac{\sum_i \theta_i \Delta F_i}{\bar{W}^\varepsilon}, \quad (10)$$

i.e. as the sum of the dollar inflows ΔF_i into each fund i , times the marginal propensity of fund i to invest in equities, θ_i , as a fraction of the the baseline value of aggregate equities $W^\varepsilon = Q\bar{P}$.²⁴ At the same time the net total flow is 0, $\sum_i \Delta F_i = 0$, as one bond removed from one fund goes to another fund, and the net amount of equities purchased is 0, $\sum_i \Delta Q_i = 0$, as the net amount of shares is constant:²⁵

$$\sum_i \Delta F_i = 0, \quad \sum_i \Delta Q_i = 0. \quad (11)$$

Hence, there is a well-defined notion of “the aggregate flow into equities,” f_S (equation (10)) which is generically non-zero, even though “for every buyer there is a seller” (equation (11)).

The impact of flows on the aggregate price We now analyze what happens after the aggregate inflow f_S in equities. We assume from now on that $\zeta > 0$. As the supply of shares does not change, we must have $q = 0$ in the equilibrium after the flow shock. Given (7), we have $0 = q = -\zeta p + f$, and the price change must be $p = \frac{f}{\zeta}$. Proposition 3 summarizes this.²⁶

²³Indeed, as $\theta_i = \frac{\bar{W}_i^\varepsilon}{\bar{W}^\varepsilon}$, we have $f_S = \sum_i S_i f_i = \sum_i \frac{\bar{W}_i^\varepsilon}{\bar{W}^\varepsilon} \frac{\Delta F_i}{\bar{W}_i} = \frac{1}{\bar{W}^\varepsilon} \sum_i \theta_i \Delta F_i^\varepsilon$.

²⁴This is analogous to the marginal propensity to take risk in Kekre and Lenel (2020).

²⁵For instance, if there are just the pure bond fund and a mixed fund, then the bond flow into the mixed fund ΔF_1 is compensated by a flow out of the pure bond fund, so $\Delta F_0 = -\Delta F_1$.

²⁶It is exact when all $\kappa_i = 0$ and it uses a first-order Taylor expansion for small flows f when $\kappa_i \neq 0$.

Proposition 3. *Suppose that the representative consumer invests ΔF_i in each fund i , so that the total inflow in equities is a fraction $f = f_S = \sum_i S_i \frac{\Delta F_i}{W_i}$ of the value of equities. Then, the stock price changes by a fraction $p := \frac{P - \bar{P}}{\bar{P}}$ equal to:*

$$p = \frac{f}{\zeta}, \quad (12)$$

where ζ is the macro elasticity of demand defined in (9).

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.²⁷

An undergraduate example To think through the economics of Proposition 3, we found the following simple, undergraduate-level example useful. Suppose that there are just two funds: the pure bond fund and the representative mixed fund, which always holds 80% in equities (the magnitude suggested by Figure 1). Then, $\theta = 0.8$, $\kappa = 0$, so that $\zeta = 1 - \theta = 0.2$ 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. We conclude with a few remarks.

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

$$f_C = \frac{\text{Net repurchases (in value)}}{\text{Total equity value}} = - \frac{\text{Net issuances (in value)}}{\text{Total equity value}}. \quad (13)$$

Then, the basic net demand for shares is as above, using the total flow:

$$f := f_S + f_C, \quad (14)$$

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,

²⁷If $d \neq 0$, there is an extra effect, and $p = \frac{f}{\zeta} + \frac{\kappa \delta}{\zeta} d$, with $\frac{\kappa \delta}{\zeta} < 1$. This implies that unaided by flows, prices under-react to fundamentals in inelastic markets.

we want to add share repurchases by corporations. In addition, if firms have a supply elasticity ζ_C , then the basic equilibrium is: $f_S - \zeta p = -f_C + \zeta_C p$. That is, a change in demand $f_S - \zeta p$ equals a change in supply $-f_C + \zeta_C p$. Therefore $p = \frac{f_S + f_C}{\zeta + \zeta_C}$, so that the effective market elasticity is $\zeta + \zeta_C$. In much of the paper, we assume that the supply of shares is inelastic, $\zeta_C = 0$, which will prove to be a good approximation.

The representative mixed fund's equity share vs. the market-wide equity share There are two notions of equity share. The traditional one is the wealth-weighted equity share:

$$\theta_W = \frac{W^\mathcal{E}}{W^\mathcal{E} + W^\mathcal{B}} = \frac{\text{Total value of Equities}}{\text{Total value of Equities + Bonds}}, \quad (15)$$

which can also be expressed as $\theta_W = \frac{\sum_i W_i \theta_i}{\sum_i W_i}$. The other one is the equity-holdings weighted equity share defined earlier, $\theta_S = \frac{\sum_i W_i^\mathcal{E} \theta_i}{\sum_i W_i^\mathcal{E}}$, where $W_i^\mathcal{E}$ was the equity holding of fund i . The former share (θ_W) is directly available in aggregated data, while the latter (θ_S) is what matters for the macro elasticity. They are different, and indeed $\theta_S > \theta_W$.²⁸ This makes the disaggregation issues potentially non-trivial, and will require some care in the empirical part.

Take the undergraduate example with just two funds, the mixed fund and the pure bond fund, and $\kappa = 0$. Then, whatever the flows, $\theta_S = \theta$ is always constant, pinned by the mandate θ of that mixed fund. However, θ_W varies over time, as flows in and out of equities change the market value of equities, P .

3.2 Infinite horizon model

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 consider the case where there is a pure bond fund and “representative mixed fund” trading stocks and bonds. This allows us to zoom in on the core economics: an economy with several funds can be represented via a single mixed fund to the leading order, as in Proposition 2.²⁹ The representative mixed fund has a mandate: the fraction invested in equities, $\frac{P_t Q_t}{W_t}$, should be

$$\frac{P_t Q_t}{W_t} = \theta e^{\kappa \hat{\pi}_t + \nu_t}, \quad (16)$$

where as before $\hat{\pi}_t := \pi_t - \bar{\pi}$ 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 shocks to tastes or perceptions of risk. We assume that dividends and interest rates on bonds are passed to consumers: hence, reinvesting dividends counts as an inflow.

To analyze this economy, it is useful to linearize it. This needs to be done around a simpler, “baseline” economy, which is on a balanced growth path with a constant equity premium $\bar{\pi}$. We call

²⁸Indeed, using $W_i^\mathcal{E} = \theta_i W_i$, $\theta_S = \mathbb{E}_S[\theta_i] = \sum_i S_i \theta_i = \frac{\sum_i W_i^\mathcal{E} \theta_i}{\sum_i W_i^\mathcal{E}} = \frac{\sum_i W_i \theta_i^2}{\sum_i W_i \theta_i} = \frac{\mathbb{E}_W[\theta_i^2]}{\mathbb{E}_W[\theta_i]} \geq \mathbb{E}_W[\theta_i] = \theta_W$. As long as there is a pure bond fund, the θ_i are not identical, and the inequality is strict. Formally, we assume that all funds have weakly positive total wealth.

²⁹This is detailed in Appendix G.7.

\bar{P}_t , \bar{D}_t , \bar{W}_t , and \bar{Q} the baseline price, dividend, wealth, and quantity of shares held by the mixed fund. We assume that $(\bar{P}_t, \bar{D}_t, \bar{W}_t) = (\bar{P}_0, \bar{D}_0, \bar{W}_0) \mathcal{G}_t$: they grow with a common cumulative growth factor \mathcal{G}_t , such that $\frac{\mathcal{G}_{t+1}}{\mathcal{G}_t}$ follows an i.i.d. growth process with mean g . As the equity premium is always $\bar{\pi}$ in the baseline economy, $r_f + \bar{\pi} - g = (1 + g)\delta$, with $\frac{\bar{P}_t \bar{Q}}{\bar{W}_t} = \theta$ and $\frac{\bar{D}_t}{\bar{P}_t} = \delta$.³⁰ At the same time, the bond holdings of the mixed fund are $\bar{B}_0 + \bar{F}_t$, where \bar{F}_t is the cumulative dollar inflow since time 0 (so $\bar{F}_0 = 0$): the only “new” bonds that the representative mixed fund has must come from inflows, like in the undergraduate model above. They should also represent a fraction $1 - \theta$ of the wealth of the fund, so that we have: $\bar{B}_0 + \bar{F}_t = \frac{1-\theta}{\theta} \bar{P}_t \bar{Q}$. This means that $\bar{F}_t = \frac{1-\theta}{\theta} (\bar{P}_t - \bar{P}_0) \bar{Q}$. This is the flow consistent with a balanced growth path in the rational economy.

We call p_t , w_t , d_t , q_t 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 $q_t = \frac{Q_t}{\bar{Q}} - 1$. We define the flow f_t as the scaled cumulative inflow in excess of the baseline:³¹

$$f_t = \frac{F_t - \bar{F}_t}{\bar{W}_t}. \quad (17)$$

We call the expected dividend deviation $d_t^e = \mathbb{E}_t d_{t+1}$. The expected excess return is $\pi_t = \frac{\mathbb{E}_t[\Delta P_{t+1} + D_{t+1}]}{\bar{P}_t} - r_f$, and we use the following Taylor expansion (see Section F for a derivation):

$$\hat{\pi}_t = \delta (d_t^e - p_t) + \mathbb{E}_t [\Delta p_{t+1}]. \quad (18)$$

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*

$$q_t = -\zeta p_t + f_t + \nu_t + \kappa \delta d_t^e + \kappa \mathbb{E}_t [\Delta p_{t+1}], \quad (19)$$

where $\zeta = 1 - \theta + \kappa \delta$ is the aggregate elasticity of the demand for stocks, as in (9).

As the total number of shares is constant, the equilibrium condition is given by $q_t = 0$. 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):*

$$p_t = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{1}{(1 + \rho)^{\tau-t+1}} \left(\rho \frac{f_\tau + \nu_\tau}{\zeta} + \delta d_\tau^e \right), \quad (20)$$

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

$$\rho = \frac{\zeta}{\kappa} = \delta + \frac{1 - \theta}{\kappa}. \quad (21)$$

The deviation of the equity premium from its average is:

$$\hat{\pi}_t = \frac{(1 - \theta) p_t - (f_t + \nu_t)}{\kappa}. \quad (22)$$

³⁰Indeed, $1 + r_f + \bar{\pi} = \mathbb{E}_t \left[\frac{\bar{P}_{t+1} + \bar{D}_{t+1}}{\bar{P}_t} \right] = \mathbb{E}_t \left[\frac{\bar{P}_{t+1}(1 + \delta)}{\bar{P}_t} \right] = (1 + g)(1 + \delta)$.

³¹This is extremely close to another definition, $f_t = \sum_{s=0}^t \frac{\Delta F_s - \Delta \bar{F}_s}{\bar{W}_{s-1}}$, but the above definition is the one warranted by the theory.

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 $\rho = \delta$.³²

In (20), the price discounts future dividends at a rate $\rho \geq \delta$ given in (21). 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)*.³³ It makes good sense that a lower sensitivity to the equity premium makes the market less reactive to the future, hence more myopic.^{34,35} In the rest of this section, we set $\nu_t = 0$; the general case simply comes from replacing f_t by $f_t + \nu_t$.

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 price at time $t \geq 0$ is (via (20), with $\mathbb{E}_0[f_\tau] = f_0$):

$$\mathbb{E}_0[p_t] = \frac{1}{\zeta} f_0. \quad (23)$$

So, the “price impact” is permanent. As a result, the equity premium is permanently lower, $\mathbb{E}_0[\hat{\pi}_t] = -\delta \frac{f_0}{\zeta}$ (see (18)) This is simply because, if the equity demand has permanently increased, equity prices should be permanently higher.³⁶

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,³⁷ 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

³²Strictly speaking, this is only true with risk-neutral arbitrageurs, so that the risk premium is 0. The general case is in Section F.4 where the elasticity is still very high.

³³The formula extends to changes in the interest rate, as in $r_{ft} = \bar{r}_f + \hat{r}_{ft}$. As (18) becomes $\hat{\pi}_t = \mathbb{E}_t \Delta p_{t+1} + \delta(d_t^e - p_t) - \hat{r}_{ft}$, all expressions are the same, replacing d_t^e by $d_t^e - \frac{1}{\delta} \hat{r}_{ft}$, including in (20). We assume here that the bond is very short term, with zero duration. If the bond has non-zero duration, there is another term corresponding to the capital gains on bonds.

³⁴The intuition for the sign of the impact of θ on ρ is as follows: 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. This myopia in (20) generates momentum: because the market is myopic (by (20)), dividend news are only slowly incorporated into the price.

³⁵Here the demand (19) depends on the equity premium as $\kappa \hat{\pi}_t = \kappa \mathbb{E}_t \Delta p_{t+1} + \kappa \delta (d_t^e - p_t)$. A variant would be 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 (20) still holds, after multiplying δd_t^e by $\frac{\kappa^D}{\kappa}$. This highlights that κ^D increases the market elasticity ζ , while κ increases market “forward-lookingness” $\frac{1}{\rho}$.

³⁶In 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 G.5 details the link with the Kyle model.

³⁷Indeed, with $p = \ln 2$, $\hat{\pi} = -\delta p = - (3\%) \times 0.7 \simeq -2\%$.

there are no further disturbances, the impact on the time- t price is $p_t = \frac{f_t}{\zeta + \kappa\phi_f}$ (see (20)), implying

$$\mathbb{E}_0 [p_t] = \frac{(1 - \phi_f)^t}{\zeta + \kappa\phi_f} f_0, \quad (24)$$

and the change in the equity premium is $\mathbb{E}_0 [\hat{\pi}_t] = -\frac{\delta + \phi_f}{\zeta + \kappa\phi_f} (1 - \phi_f)^t f_0$ (see (22)). 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 on impact (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 $p_t = \frac{1}{(1+\rho)^{T-t}} \frac{f_T}{\zeta}$ (see (20), using $f_\tau = 1_{\tau \geq T} f_T$), so that after the initial jump, the price gradually drifts upward (assuming for concreteness that the inflow is positive). Hence, the risk premium is elevated by $\hat{\pi}_t = \frac{1-\theta}{\kappa} p_t$ (for $t \in [0, T]$, see (22)), and more elevated as one nears the inflow. After the inflow, though, we are back to the case of a permanently elevated price and permanently 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.

A simple benchmark To think about the stochastic steady state, it is useful to consider f_t as an autoregressive process with speed of mean-reversion ϕ_f :

$$f_t = (1 - \phi_f) f_{t-1} + \varepsilon_t^f, \quad (25)$$

with $\mathbb{E}_{t-1} [\varepsilon_t^f] = 0$. Then, a high inflow increases equity prices and hence lowers the equity premium, in the following precise manner:³⁸

$$p_t = b_f^p f_t, \quad \hat{\pi}_t = b_f^\pi f_t, \quad b_f^p = \frac{1}{\zeta + \kappa\phi_f}, \quad b_f^\pi = -(\delta + \phi_f) b_f^p. \quad (26)$$

Calibration We want to understand how a macro price impact of $M \simeq 5$ might arise, and for this we calibrate the model. When flows are mean-reverting with speed ϕ_f , the price impact is $M = \frac{1}{\zeta^M}$, with $\zeta^M = \zeta + \kappa\phi_f = 1 - \theta + \kappa(\delta + \phi_f)$ (see (9), (24), and (26)). Some parameters are easy to estimate. We take a dividend-price ratio $\delta = \frac{D}{P} = 3.7\%$ /year (we use annualized units throughout).³⁹ We calibrate $\phi_f = 4\%$ /year to match the speed of mean-reversion of the dividend-price ratio.⁴⁰ Given the results in Figure 1, we take an equity share $\theta = 87.5\%$ (equity-holdings weighted as in θ_S).

Calibrating κ is most challenging. 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

³⁸This can be derived by plugging in those values in (19) with $q = 0$ in equilibrium, or via (20).

³⁹Section H.1 details how to go from continuous to discrete time.

⁴⁰We 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 $\frac{1+3\rho}{T}$ over our sample of $T = 72$ years, which is around $\frac{4}{72}$. Thus we calibrate $\phi_f = 1 - \rho^{OLS} - \frac{4}{72} \simeq 4\%$.

$\kappa = \frac{d \ln \theta_{it}}{d \pi_t} = \frac{1}{\pi} = 22$, using an annual equity premium of 4.4%.⁴¹ But, we rarely 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 $\pi_t = 5\%$ to $\pi_t = 10\%$, 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 $\kappa_i = \frac{d \ln \theta_i}{d \pi} = \frac{\ln 0.75 - \ln 0.5}{0.05} \simeq 8$. 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_i for a typical fund with equity share of 50% equal to about 4. Moreover, a 100% equity funds needs to have $\kappa_i = 0$; 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 $K \simeq 8$. So, we obtain $\kappa = \kappa_S = K(1 - \theta_S) \simeq 8 \times (1 - 0.88) \simeq 1$. This gives a simple microeconomic interpretation for the value $\kappa = 1$. Together, this yields $\zeta = 0.16$, and $\zeta^M = 0.2$, so that the price impact is indeed $M = \frac{1}{\zeta^M} = 5$. If the flows are extremely persistent, the subtle difference between ζ and ζ^M vanishes ($\kappa \phi_f$, which is 0.04 in the calibration, goes to 0).

4 Estimating the Aggregate Market Multiplier and Elasticities

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 the macro elasticity is a new parameter of interest that is first estimated in this paper.⁴²

The key difficulty is that prices, equity demand, and flows are in part driven by aggregate shocks, such as macroeconomic news, so that naively regressing prices on flows or flows on prices would not yield a consistent estimate of the elasticity or multiplier. Hence, we need an instrument.

We show how to estimate macro multipliers and elasticities by decomposing demand shocks of different investors into (i) aggregate shocks, which shift the equity demand of all investors at the same time (albeit with potentially different loadings), and (ii) idiosyncratic shocks that only shift the equity demand of a subset of investors. We first develop the general intuition in Section 4.1. We also use this framework to interpret the identifying assumptions in Table 1 that have been used in the related literature. Our empirical strategy is an application of the method called Granular Instrumental Variables (GIV), which we conceived for the present paper, and lay out in Gabaix and Koijen (2020).

In Section 4.2 to Section 4.3, we provide two implementations of our general idea. First, in Section 4.2, we estimate the aggregate multiplier using shocks to mutual fund flows and data from 13F filings to construct aggregate factors. Second, in Section 4.3, we use data from the Flow of Funds to estimate the aggregate elasticity and multiplier.

⁴¹This is for a fund maximizing rationally a CRRA function of financial wealth. In Section F.4 we consider a more sophisticated thought experiment, with a consumer maximizing lifetime utility out of labor income in addition to financial wealth. Then, the value of ζ is even higher.

⁴²See Table 1 for a summary of related estimates in the literature.

Each settings has advantages and disadvantages in terms of the assumptions that we have to make, and there are differences in the quality of the data that are available to us in estimating the elasticities and the multiplier. With that said, the combined evidence suggests that the aggregate stock market is inelastic (in line with the findings of Table 1).

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.

We conclude this section by connecting capital flows to macroeconomic variables and measures of beliefs to provide an initial analysis of the potential determinants of flows into the equity market.

4.1 Estimating aggregate elasticities and multipliers

Notation We first provide a brief summary of the GIV method—the appendix and Gabaix and Koijen (2020) provide further details, such as a justification of consistency and extensions. Recall that we use the following notations, with the shares S_i adding up to 1:

$$X_{Et} := \frac{1}{N} \sum_{i=1}^N X_{it}, \quad X_{St} := \sum_{i=1}^N S_{i,t-1} X_{it}, \quad X_{St}^{-i} := \sum_{j,j \neq i} S_{j,t-1} X_{jt}, \quad X_{\Gamma t} := X_{St} - X_{Et}. \quad (27)$$

The shares, $S_{it} = \frac{Q_{it}}{\sum_i Q_{it}}$, is the fraction of the market held by investor i .

A factor-based demand system for the aggregate stock market Let $\Delta q_{it} = \frac{Q_{it} - Q_{i,t-1}}{Q_{i,t-1}}$ denote a time series of fractional changes in investors' equity holdings, where i indexes investors as before. To develop the main idea, we model Δq_{it} as (omitting constants):⁴³

$$\Delta q_{it} = -\zeta_i \Delta p_t + f_{it}^\nu, \quad (28)$$

where Δp_t is the aggregate stock return and ζ_i is the demand elasticity of investor i . Depending on the empirical setting, we will impose additional restrictions on the elasticities.

We model the demand disturbance f_{it}^ν as:

$$f_{it}^\nu = \lambda_i' \eta_t + u_{it}, \quad (29)$$

where η_t is a vector of common shocks (which can include observable factors, such as GDP growth, or latent factors), λ_i is a vector of factor loadings, and u_{it} is an idiosyncratic shock. The key identifying assumption that we make throughout is that

$$\mathbb{E}[u_{it} \eta_t] = 0. \quad (30)$$

Estimating the multiplier and elasticities Aggregating the individuals demands (28) as in Proposition 2, the aggregate demand is $\Delta q_{St} = -\zeta_S \Delta p_t + f_{St}^\nu$. Using market clearing, we have $\Delta q_{St} = 0$, that is

$$\Delta p_t = M (\lambda_S' \eta_t + u_{St}), \quad (31)$$

for the multiplier

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

⁴³To lighten things up, we simplify a bit the notations. Compared to (19), we use the notation f_{it}^ν for $\Delta f_{it}^\nu := \Delta f_{it} + \Delta \nu_{it}$, whete we absorb the change-in-expectation terms $\kappa_i \Delta \mathbb{E}_t [\delta d_t^e + \Delta p_{t+1}]$ into the “demand shifter” $\Delta \nu_{it}$.

The goal is to estimate the multiplier M or, equivalently, the aggregate elasticity, ζ_S .

If we substitute the market clearing price into the demand equation (28), we have

$$\Delta q_{it} = \tilde{\lambda}'_i \eta_t - \frac{\zeta_i}{\zeta_S} u_{St} + u_{it}, \quad (32)$$

where $\tilde{\lambda}_i = \lambda_i - \frac{\zeta_i}{\zeta_S} \lambda_S$.

The key idea of the GIV method is to identify M and ζ_S using variation that comes from the idiosyncratic shocks, u_{it} . Indeed, if we have access to u_{St} , we can just estimate M by OLS, regressing $\Delta p_t = M u_{St} + \varepsilon_t$ (this is warranted from our assumption (30)). Likewise, to estimate the elasticity ζ_i of a sector i , we can use the (size-weighted average) of idiosyncratic shocks u_{-it} across other sectors to instrument the price. The GIV method also provides a way to estimate the idiosyncratic shocks u_{it} , or a proxy for them: we explain the specifics below.

While we can use the idiosyncratic shocks u_{it} , it is worth highlighting that we cannot use aggregate shocks η_t for identification. As can be seen from (32), the covariance matrix of demand changes can identify $\tilde{\lambda}_i$, which depends on λ_i and $\frac{\zeta_i}{\zeta_S}$, hence we cannot identify our parameters of interest. Intuitively, if an investor sells in response to an aggregate shock, we do not know whether this is because of a high exposure in λ_i or because the investor’s demand elasticity differs from the aggregate demand elasticity, $\zeta_i \neq \zeta_S$.

We detail in each application how to estimate those idiosyncratic shocks empirically to make the above reasoning valid. The general case is covered in Gabaix and Koijen (2020).

GIV: Requirements and threats to identification For the GIV to be consistent, we need $\mathbb{E}[u_{it}\eta_t] = 0$ 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 sectors, so that the market is “granular” in the sense that the idiosyncratic shocks to a few large sectors meaningfully affect the aggregate.⁴⁴ 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)). Ben-David et al. (forthcoming) and Ghysels et al. (2021) study the impact of investor granularity on the cross-section of US stock returns.

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.

Connection to other identifying strategies in the literature The intuition behind our identification strategy ties back to the empirical strategies that have been used in the literature so far. Indeed, the typical approach is to isolate an idiosyncratic demand shock to a (group of) investors that does not affect the demand of other investors other than through price.

The classic approach is to look at stocks that have been included in (or deleted from) a benchmark. If changes in the composition of the benchmark are random, the inelastic demand from

⁴⁴Indeed, when flow shocks have volatility σ_u , $\text{var}(u_S) = H\sigma_u^2$, 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.

benchmark-restricted investors provide exogenous demand shocks. This demand shock is idiosyncratic from the perspective of investors who are not compared to the benchmark. This u_{it} from benchmark-restricted investors can thus be used to estimate the (micro) multiplier and the elasticity of unconstrained investors.

As a second example, a recent literature that constructs instruments using payouts by firms. The basic intuition is that there is a long lag between the date that a dividend is declared (and thus when investors learn new information about the firm or the market) and when the dividend is actually paid. This anticipated cash flow provides a shift in demand and several papers have used it to estimate micro and macro multipliers (Schmickler, 2020; Hartzmark and Solomon, 2022).

As a third example, Ben-David et al. (2020a) use a change in the Morningstar ratings methodology that changed how flows are directed to different factors. This is an idiosyncratic demand shocks from the perspective of all investors other than mutual funds and can thus be used as a valid instrument.

Obtaining such events at the level of the aggregate stock market is challenging, and we therefore rely on estimating a factor model to isolate the idiosyncratic shocks. In each of the three applications, we discuss how we specialize the general methodology to our specific setting.

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 3.1. When flows mean-revert with speed ϕ_f , the measured elasticity is $\zeta + \kappa\phi_f$, as we saw in (24) and (26). Combining those two extensions, the measured price impact is

$$M = \frac{1}{\zeta^M}, \quad \zeta^M := \zeta + \kappa\phi_f + \zeta_C. \quad (33)$$

As $\kappa\phi_f$ and ζ_C appear to be small, the difference between ζ , $\zeta + \kappa\phi_f$ and ζ^M is rather minor, and is best ignored in the first pass. Still, to be completely explicit, when we empirically measure “ ζ ,” we actually measure a quantity that is $\zeta + \kappa\phi_f + \zeta_C$ if flows mean-revert at speed ϕ_f and firms have a supply elasticity ζ_C , and is strictly speaking ζ only when flows do not mean-revert and $\zeta_C = 0$.

4.2 Multiplier estimates: Evidence from mutual fund flows

Data In this section, we combine investor-level data from 13F filings and fund-level mutual fund data.⁴⁵ We use 13F data from FactSet that cover the period from 2000.Q1 to 2019.Q4, and we follow the data construction as in Kojien et al. (2019). Monthly mutual fund data come from Morningstar and it covers the period from January 1993 to December 2019.

We use the disaggregated mutual fund data to construct a measure of aggregate mutual fund flows. We use equity weights in aggregating the flows of individual funds, as warranted by our theory, see (10). Concretely, we start from the share invested in US equities by fund i , θ_{it} , assets under management, W_{it} , and the flow ΔF_{it} as defined by Morningstar. We compute $\Delta f_{it} = \frac{\Delta F_{it}}{W_{i,t-1}}$ and measure the aggregate flow into mutual funds as $\Delta f_t^{MF} = \frac{\sum_i \theta_{i,t-1} W_{i,t-1} \Delta f_{it}}{\sum_i \theta_{i,t-1} W_{i,t-1}}$, which uses equity-weighting. We include “US equity,” “sector equity,” “international equity,” and “allocation” funds in the analysis. When equity shares are missing at a monthly frequency, which are only used in determining the weighting scheme, we fill them in using the most recent value for a given fund.

⁴⁵In 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.

Lastly, we winsorize Δf_{it} at 1% and 99% in each period before computing the aggregate flow. We provide further details on the data construction in Appendix C.2.

Methodology Our first estimate of the multiplier starts from equation (31). We decompose $u_{St} = u_{St}^{-MF} + S_{t-1}^{MF} u_t^{MF}$ and write the price equation as

$$\Delta p_t = M \lambda'_S \eta_t + M u_{St}^{-MF} + M S_{t-1}^{MF} u_t^{MF}. \quad (34)$$

The common factors, η_t , capture the comovement between the flows of mutual fund investors and the aggregate demand shocks of other investors, for instance, in response to macroeconomic news. The second term, u_{St}^{-MF} , captures the idiosyncratic demand shocks of all other investors. The final term captures the shocks specific to the mutual fund sector, $S_{t-1}^{MF} u_t^{MF}$, multiplied by M . Our objective in this section is to measure $S_{t-1}^{MF} u_t^{MF}$ and then use it to estimate the multiplier, M .

We proceed in four steps:

1. We estimate innovations in mutual fund flows by removing the predictable component:

$$\Delta f_t^{MF} = a_0 + \sum_{l=1}^k a_l \Delta f_{t-l}^{MF} + ct + \epsilon_{mt}^{MF}, \quad (35)$$

We estimate the model at a monthly frequency. The estimates are reported in Table D.12 in Online Appendix D.6. Based on the estimated model $AR(k)$ model for flows, we define $K = \frac{1}{1 - \sum_{l=1}^k a_l}$. It is the cumulative flow due to shocks ϵ_{mt}^{MF} . $K \epsilon_{mt}^{MF}$ is what matters according to Proposition 5 as it measures the cumulative flow following an innovation ϵ_{mt}^{MF} .⁴⁶ We aggregate the monthly innovations, ϵ_{mt}^{MF} , for $k = 3$, in each quarter and refer to these innovations as ϵ_t^{MF} .

2. It is well known that the innovations, ϵ_t^{MF} , are correlated with contemporaneous realized returns (Warther (1995); Edelen and Warner (2001); Goetzmann and Massa (2003)). However, this correlation may simply be driven by aggregate demand shocks, η_t . We extend this literature by removing aggregate demand factors and isolating the idiosyncratic demand shocks of mutual fund investors. Concretely, we model

$$\epsilon_t^{MF} = \beta'_0 \eta_t + \beta'_1 C_t + u_t^{MF}, \quad (36)$$

where η_t are common unobserved factors, C_t are common observed factors, and u_t^{MF} are the unique shocks to fund flows.

3. To estimate η_t , we use the 13F filings of investors outside of the mutual fund industry (e.g., pension funds, insurance companies, and so forth). We omit investors in the mutual fund industry, using the same assignment of investor types as in Koijen et al. (2019), by removing investment advisors and mutual funds. To estimate η_t , we consider an extension of the model in (28) and allow for heterogeneity in demand elasticities, $\zeta_{i,t-1}$, across investors and over time:

$$\Delta q_{it} = \alpha_i - \zeta_{i,t-1} \Delta p_t + \lambda'_{i,t-1} \eta_t + \beta'_i C_t + u_{it}.$$

⁴⁶We abstract from additional discounting at a rate ρ , as it is immaterial at the horizon of a few months that we use. See Section G.5 for details.

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

$$\zeta_{i,t-1} = \dot{\zeta}' x_{i,t-1}, \quad \lambda_{i,t-1} = \dot{\lambda}' x_{i,t-1} + \ddot{\lambda}_i,$$

where $x_{i,t-1}$ is a vector of investor characteristics of which the first element is equal to 1, and $\dot{\zeta}$, $\dot{\lambda}$, and $\ddot{\lambda}_i$ are to be estimated. As investor characteristics, we use an investor's active share and log AUM. In addition, we allow for non-parametric factors via $\ddot{\lambda}_i$. We also control for macroeconomic factors, C_t , and allow for heterogeneous exposures via $\beta_i' C_t$. We discuss in Section B.1 how we estimate the common factors, η_t , and we refer to the estimates as η_t^e .

4. With estimates of ϵ_t^{MF} and η_t in hand, we regress returns on fund flow innovations, while controlling for common factors,

$$\Delta p_t = a + M Z_t + \lambda' \eta_t^e + m' C_t + e_t,$$

where $Z_t = K S_{t-1}^{MF} \epsilon_t^{MF}$. After controls, Z_t is the surprise inflow unique to mutual funds, and implements (34). As a common observed factor, C_t , we use GDP growth. We also explore robustness to controlling for changes in volatility in C_t .

Empirical results The results are summarized in Table 2. The first column presents the results with only Z_t and GDP growth. The next four columns add the factors extracted from the 13F data, η_t^e , as recommended by the GIV. In the final column, we also control for the quarterly (percentage) change in volatility. Without controls other than GDP in Column 1, the multiplier estimate equals $M = 10.9$. By adding additional controls, the R-squared value increases significantly and the multiplier estimate lowers, as we would expect since demand shocks and prices are positively correlated. With four additional factors, the R-squared value equals approximately 60% and the multiplier drops to $M = 7.7$ with a standard error of 2.3. In the final column, we add changes in volatility. While these do not correlate strongly with fund flow innovations, they do correlate with returns. This suggests that other investors are negatively sensitive to volatility and this also captures a source of demand shocks. The multiplier lowers further to 7.6 and the R-squared is now 70%.⁴⁷ In Figure D.8, we also repeat the long-horizon analysis as in Figure 4. As before the impact of flow shocks on prices is persistent although the confidence interval is wide at longer horizons of one year.

In summary, we find that the multiplier estimates are quite consistent with the estimates we found using the FoF data. These estimates well above 1 are consistent with the estimates for other countries and for style factors, see Table 1. Future research can explore other strategies to control for common demand factors to sharpen the identification.

4.3 Multiplier and elasticity estimates: Evidence from the Flow of Funds

For our second application, we use data from the Flow of Funds. The advantages are that we use a longer sample and that we can use the idiosyncratic shocks of all sectors, not just mutual funds. As such, it is a typical implementation of the GIV procedure.

A disadvantage is that the Flow of Funds data require additional measurement assumptions as we discuss in Online Appendix C. Also, as the number of sectors is relatively small, we need to make an additional assumption that $\zeta_i = \zeta$.

⁴⁷As volatility is endogenous, it can be included only with interpretative circumspection. We include it here for descriptive purposes.

Table 2: Estimates of the macro elasticity using mutual fund and 13F data. The first five columns provide estimates of the multiplier M , which is the coefficient on $Z_t = KS_{t-1}^{MF}\epsilon_t^f$, the innovation in the cumulated inflow into mutual funds after controls. We regress returns on unexpected flows, ϵ_t^f , times the share of aggregate equities held by the mutual fund sector, S_{t-1}^{MF} , and adjusting for the fact that inflows are autocorrelated (see (35) and the surrounding definition of K). In the first column we only control for GDP growth and in the next four columns we add one to four common factors to isolate the idiosyncratic component in mutual fund flows. The common factors are extracted from 13F filings of institutions outside of the mutual fund industry. In the final column, we add the change in quarterly volatility as an additional control. We report the standard errors, which are corrected for autocorrelation, in parentheses. The sample is from 2000.Q1 to 2019.Q4.

	Δp					
Z	10.93 (2.64)	10.85 (2.78)	8.54 (2.18)	7.69 (2.32)	7.69 (2.34)	7.62 (1.92)
GDP growth	4.19 (1.23)	4.21 (1.25)	4.94 (0.96)	4.99 (1.17)	5.00 (1.17)	3.43 (1.17)
PC1		-0.91 (0.74)	-0.94 (0.60)	-0.95 (0.63)	-0.95 (0.63)	0.06 (0.61)
PC2			-2.98 (0.66)	-3.07 (0.48)	-3.07 (0.48)	-0.82 (0.37)
PC3				-0.87 (0.56)	-0.87 (0.56)	-1.25 (0.41)
PC4					0.11 (0.33)	-0.21 (0.38)
$\Delta\sigma$						-0.10 (0.01)
Constant	0.00 (0.00)	0.00 (0.00)	-0.00 (0.01)	-0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
Observations	80	80	80	80	80	80
R^2	0.426	0.438	0.556	0.565	0.566	0.699

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

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

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:

$$Z_t = u_{St} - u_{Et} = u_{\Gamma t}. \quad (37)$$

As $u_{\Gamma t}$ is a combination of idiosyncratic shocks only, it is uncorrelated with η_t , see (30). This orthogonality condition makes $Z_t = u_{\Gamma t}$ a valid instrument: it is our GIV. Furthermore, if u_{it} is homoskedastic, then $u_{\Gamma t}$ is uncorrelated with u_{Et} .⁴⁸ 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

$$\Delta p_t = M Z_t + e_t, \quad (38)$$

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

$$\Delta q_{Et} = -\zeta \Delta p_t + \epsilon_t, \quad (39)$$

with Δp_t instrumented by Z_t .

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. Viewed this way, the GIV estimator generalizes the idea behind the index inclusion 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.

We reiterate that the methodology works even if we do not have data on flows f_{it} —it is enough to have data on changes in equity holdings Δq_{it} . This implies that we identify idiosyncratic shocks to $f'_{it} = f_{it} + \nu_{it}$, where f_{it} are capital flows and ν_{it} are demand shocks.

Methodology: 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 $\check{a}_{it} := a_{it} - a_{Et}$, that is, the cross-sectionally demeaned value of a vector a_t . We run a principal component analysis (PCA) via the model

$$\Delta \check{q}_{it} = \check{\lambda}'_i \eta_t + \check{u}_{it}. \quad (40)$$

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:

$$\Delta p_t = M Z_t + \beta' \eta_t + e_t, \quad (41)$$

and estimate 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 its optimality by various

⁴⁸The same condition holds in the more general case of uncorrelated heteroskedastic u_{it} , with the inverse variance weights \bar{E} , so $Z_t := \Delta q_{St} - \Delta q_{\bar{E}t}$ (see Gabaix and Koijen (2020)).

metrics (e.g. it is GMM optimal). 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.2. We demonstrate the performance of the GIV estimator in our specific setting in Appendix D.1 using simulations.

Empirical results 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

$$\Delta q_{it} = \alpha_i - \zeta \Delta p_t + \lambda_i' \eta_t + u_{it}, \quad (42)$$

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,⁴⁹ and one or more latent factors, η_t^{PC} .

The results are presented in Table 3. 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.

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 1.9.⁵⁰ 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 (39). 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 ζ_C of the corporate sector. The short-run elasticity is low at $\zeta_C = 0.01$ for both one and two principal components.⁵¹ 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.

⁴⁹We 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.

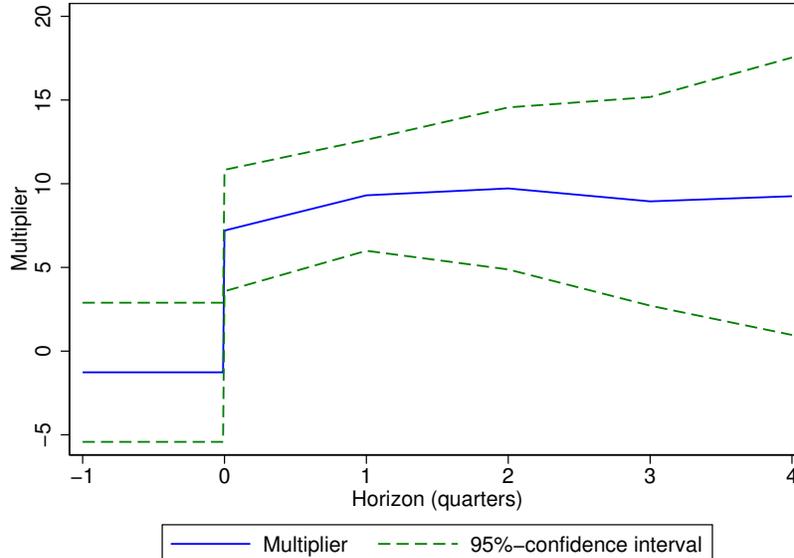
⁵⁰We report Newey-West standard errors using the bandwidth selection as in Newey and West (1994).

⁵¹This 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 (18)) $\zeta_C = \delta \times 0.06 = 0.0024$. 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.

Table 3: 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 $Z_t = \sum_i S_{i,t-1} \Delta \check{q}_{it}$ and $\Delta \check{q}_{it}$ defined in (67), to estimate the impact of sector-specific shocks on prices. In constructing $\Delta \check{q}_{it}$, 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.

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

Figure 4: 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 (43). 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.



The impact of flows at longer horizons In Figure 4, we explore how demand and flow shocks propagate across time. To this end, we extend the earlier analysis by estimating

$$p_{t+h} - p_{t-1} = a_h + M_h Z_t + c_h \eta_t^{PC,e} + d_h \Delta y_t + \epsilon_{t+h}, \quad (43)$$

for $h = 0, 1, \dots, 4$ quarters, where $p_{t+h} - p_{t-1}$ is the $(h + 1)$ -quarter (geometric) return on the aggregate stock market. Recall that $\eta_t^{PC,e}$ is the principal component, extracted in the third step of the GIV procedure as outlined in Section B.2. 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 fairly 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.4. 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.

4.4 A new measure of capital flows into the stock market

In this final section, we construct a new measure of capital flows into the stock market consistent with the theory in Section 3. While our theory provides conceptual clarity in terms of how to measure flows into the market, and to get around the problem that “for every buyer there is a seller,” the data required are unfortunately not available for all investors.

In Section 4.4, we propose a way to construct an empirical counterpart to the measure based on the available data. As this measure is new to the literature, we show its connection to prices, macroeconomic variables, and beliefs. The results in this section provide an initial analysis of the potential determinants of flows into the aggregate stock market. These results are intentionally descriptive in nature and understanding the primitive drivers of these flows is an important task for future research.

Measuring flows into the stock market We rely on the FoF data for these calculations and we refer to Appendix C for details on the data construction of the fixed income positions and flows. As (10) shows, the flow into the aggregate stock market can be measured by first computing the flow for each investor, $\Delta f_{it} = \frac{\Delta F_{it}}{W_{i,t-1}}$, and then computing the equity weighted average, $\Delta f_{St} = \sum_i \frac{\theta_{i,t-1} W_{i,t-1}}{\sum_i \theta_{i,t-1} W_{i,t-1}} \Delta f_{it}$. Unfortunately, the FoF aggregates data across many institutions and the reported flows can be mismeasured by this definition. 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. 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

$$\Delta q_{it} = \alpha + \beta_i f_{it} + \gamma_i \Delta p_t + \delta_i \Delta y_t + \epsilon_{it}. \quad (44)$$

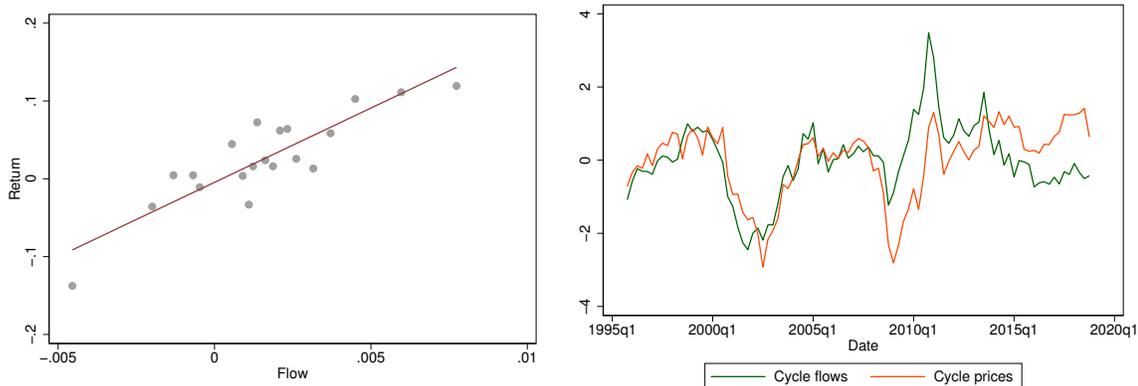
We report the estimates of β_j in Table D.13 in Appendix D. When we cannot reject $H_0 : \beta_i = 1$ at the 5% significance level, we use the total flow. If this null hypothesis is rejected, we use the equity flow instead. We refer to these “screened flows” as f_{it}^* . The aggregate flow is then given by $f_{St}^* = \sum_i S_{i,t-1} f_{it}^* + f_{Ct}$, where f_{Ct} corresponds to net repurchases of equities by firms.

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 5, 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, the slope is not a good measure of the impact of flows of the price. This is why earlier we developed an IV strategy to measure that impact.⁵²

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, by removing the time-series mean and dividing them by their standard

⁵²If 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 5: 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). In the right panel, we standardize both measures over the full sample to be able to plot them in the same figure so that they have mean 0 and standard deviation 1. The sample for both panels is from 1993.Q1 to 2018.Q4.



deviations, over the full sample to be able to plot them in the same figure. These are shown in the right panel of Figure 5. Consistent with the high-frequency co-movement that we uncover in the left panel of Figure 5, 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 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 (except returns). We standardize each of the innovation series, by removing the time-series mean and dividing them by their standard deviations, to simplify the interpretations of the regressions.

The results are presented in Table 4. 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 future 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. (2021a). 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

Table 4: 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 (0.10)		0.46 (0.11)		0.61 (0.09)		0.33 (0.09)
GDP growth		0.21 (0.11)	0.06 (0.11)			0.41 (0.10)	0.21 (0.08)
Flow				0.65 (0.09)			0.45 (0.09)
Constant	0.00 (0.10)	0.00 (0.11)	0.00 (0.10)	0.00 (0.09)	0.00 (0.09)	0.00 (0.10)	0.00 (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

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

For simplicity, we discuss in detail an endowment economy. It will be easy to then generalize the model to a production economy. This general equilibrium model is a specialization of our infinite-horizon model of Section 3.2 – it specifies things left general in that model, such as the origin of the interest rate.

The endowment Y_t follows a proportional growth process, with an i.i.d. lognormal growth rate G_t : $\frac{Y_t}{Y_{t-1}} = G_t = e^{g+\varepsilon_t^y - \frac{1}{2}\sigma_y^2}$, with $\varepsilon_{t+1}^y \sim \mathcal{N}(0, \sigma_y^2)$. Utility is $\sum_t \beta^t u(C_t)$ with $u(C) = \frac{C^{1-\gamma}}{1-\gamma}$. Because empirically dividend growth and GDP growth 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, $\frac{\mathcal{D}_t}{\mathcal{D}_{t-1}} = G_t^D = e^{g+\varepsilon_t^D - \frac{1}{2}\sigma_D^2}$, so that the balanced growth path specified in Section 3.2 has a cumulative growth factor $\mathcal{G}_t = G_t^D \dots G_1^D$. The “residual” Ω_t can be thought of as a combination of wages, entrepreneurial income, and so forth (and indeed it is the vast majority of GDP).⁵³ The representative firm raises capital entirely through equity, and passes the endowment stream as a per-share dividend $D_t = \frac{\mathcal{D}_t}{Q}$, where Q 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.⁵⁴ We write the price of equities as $P_t = \frac{D_t}{\delta} e^{p_t}$, where δ is the average dividend-price ratio and p_t 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 representative mixed fund, which holds bonds and equities. The mixed fund has a mandate, to hold a fraction in equities equal to:

$$\theta_t = \theta \exp(-\kappa^D p_t + \kappa \mathbb{E}_t[\Delta p_{t+1}]), \quad (45)$$

which is the same as before in (1), to the leading order (in terms of deviations from the steady state), with $\kappa^D = \kappa\delta$. The formulation here is slightly more general.⁵⁵

Consumption and investment by households We describe the behavior of the representative household. Section G.8 provides more formalism and further details. The dynamic budget constraint of household h entails:⁵⁶

$$Q_t^{B,h} + D_t^h + \Omega_t^h = C_t^h + \Delta F_t^h + \frac{Q_{t+1}^{B,h}}{R_{f,t}}. \quad (46)$$

Indeed, the left-hand side is the bond asset position of the household 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

⁵³Formally, it could become negative, as in Campbell and Cochrane (1999), 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.

⁵⁴We can easily have the government issue bonds, backed by taxation, see Section G.8.1.

⁵⁵But 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 (Kojien et al. (2018)), so having two κ ’s is sensible.

⁵⁶There is also the usual transversality condition, $\lim_{t \rightarrow \infty} \beta^t (C_t^h)^{-\gamma} Q_t^{B,h} = 0$.

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, 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.

The rational consumer part of the household chooses consumption (but not equity shares) to maximize lifetime utility, subject to the dynamic budget constraint for bonds (46). She takes the actions of the investor as given.⁵⁷ As she is rational, she satisfies the Euler equation for bonds:

$$\mathbb{E}_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{ft} \right] = 1, \quad (47)$$

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 (as in Barberis et al. (2006)). He seeks to maximize $\mathbb{E}_t [V^p(W_{t+1})]$ with $V^p(W) = \frac{W^{1-\gamma}-1}{1-\gamma}$ a proxy value function. Specifically, when $b_t = 0$, he chooses his allocation $\bar{\theta}^M$ in the mixed fund as:

$$\bar{\theta}^M = \operatorname{argmax}_{\theta^M} \mathbb{E} [V^p((1 - \theta^M) R_{ft} + \theta^M R_{M,t+1}) | b_t = 0], \quad (48)$$

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 fully rational value function, which would have all the Merton-style hedging demand terms, and would lead to the consumption CAPM holding on average: in particular, the equity premium would be too small, as in the equity premium puzzle (at $\bar{\pi} = \gamma \operatorname{cov}(\varepsilon_t^D, \varepsilon_t^y)$). Instead, the above formulation with narrow framing will lead to a high equity premium $\bar{\pi} = \gamma \sigma_r^2$, where the σ_r^2 is the volatility of the stock market, which is affected by flow shocks.⁵⁸

If there are no behavioral disturbances, an investor wishing to maintain a constant allocation $\bar{\theta}^M$ in the mixed fund should invest via $\bar{F}_t = \frac{1-\theta}{\theta} (\bar{P}_t - \bar{P}_0) \bar{Q}$, as in Section 3.2, that is, $\Delta \bar{F}_t = \frac{1-\theta}{\theta \delta} \Delta \mathcal{D}_t$. We assume that his policy, however, is affected by the behavioral disturbance b_t , so that the actual flow is

$$\Delta F_t = \frac{1-\theta}{\theta \delta} \Delta \mathcal{D}_t + \frac{1}{\delta} \Delta (b_t \mathcal{D}_t), \quad (49)$$

which is higher than the baseline amount $\Delta \bar{F}_t$ by a fraction Δb_t of the “fundamental value” $\frac{\mathcal{D}_t}{\delta}$ of the equity market. Here we will specify that b_t is an AR(1).

In Appendix G.9, we provide a formal microfoundation of flows via beliefs: the financier part of the household believes that the deviation of the equity premium from trend is $\hat{\pi}_t^H$. Under simple conditions, this leads to a flow

$$f_t = \kappa^H \hat{\pi}_t^H, \quad (50)$$

⁵⁷One could imagine a variant, where the consumer manipulates the investor’s actions. This would lead her to distort her Euler equation for consumption.

⁵⁸This choice of “narrow framing” leads to a high equity premium. It could be replaced by another device such as disasters. We choose here narrow framing as this behavioral ingredient is well in the behavioral spirit of this section.

with κ^H the sensitivity to the risk premium, and to a behavioral deviation $b_t = \frac{f_t}{\theta}$. Using the empirical findings of Giglio et al. (2021a), we estimate that $\kappa^H \simeq 2$, a value that we rationalize by calibrating it in terms of other behavioral parameters. This estimate is in contrast with a rational model, which would imply $\kappa^H = \frac{1}{\pi} \simeq 22$, a very large pass-through from beliefs to portfolio shares. A low value of κ^H means that people have “bold forecasts” (excess variations in the perceived equity premium) but make “timid choices” (small flows), very much as in Kahneman and Lovallo (1993).⁵⁹

This type of model can be also made to match the perspective in Bordalo et al. (2020), in which all variation in prices, flows, and the perceived risk premium $\hat{\pi}_t^H$ comes from changes in the long-term growth forecast g_t (all in deviations from a trend), in a way still governed by (50): Section G.9 provides details and a calibration. One could image a richer model for the perceived risk premium $\hat{\pi}_t^H$, e.g. with extrapolative beliefs based on realized returns or growth rates. One could then work out the implications for flows (via (50)) and prices (via Proposition (5)).

We conclude that linking flows to beliefs is a promising and manageable line of research, and the analytics that we provide in this section and in Appendix G.9 help thinking about this. At the same time, there may be other determinants of flows, for instance binding risk constraints, changes in regulation or policy, and reaction to fairly irrelevant news, which is why we find it useful to separate the impact of the behavioral deviation b_t from its determinants.

We finally formally define the equilibrium.

Definition 1. The state vector is $Z_t = (Y_t, \mathcal{D}_t, \mathcal{D}_{t-1}, b_t)$. 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: (i) the consumer follows the consumption policy $C(Z)$, which maximizes utility subject to the above constraints; (ii) the investor follows the behavioral policy (49), where the average allocation in the mixed fund is given by (48), so that it is quasi-rational with narrow framing on average, but with disturbance b_t ; (iii) the mixed fund’s demand for stocks $Q(Z)$ follows its mandate (45); (iv) the consumption market clears, $C(Z) = Y(Z)$; and (v) the equity market clears, $Q(Z) = Q$.

5.2 Model solution

Proposition 6 describes the solution of this economy. In particular, it shows that the link between the disturbance b_t and the cumulative flow f_t is as follows. Starting from an equilibrium situation, where $b_0 = 0$, the cumulative “excess” flow is equal to:

$$f_t = \theta b_t. \quad (51)$$

This holds for any process b_t . Now, we specialize to the case where b_t follows an AR(1) with speed of mean-reversion ϕ_f . Then, so does f_t , so that we are in the “simple benchmark” case of (25)-(26), and now with an endogenous interest rate and unconditional equity premium. 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 defer to future research for richer, empirically-grounded models of the “behavioral deviation” b_t , and hence of the flows. The limited goal of this framework is to have a simple model of the *impact* of the flows in general equilibrium, which can be fully solved and which lends itself to a number of variants. Importantly, it relies on observable flows.

⁵⁹Quantitatively, to match the calibrated volatility of flows of $\sigma_f = 2.8\%$ (as in Table 6) we need a moderate variation of beliefs $\sigma_{\pi^H} = 1.4\%$.

Proposition 6. *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 f_t . The market elasticity ζ and the “macro market effective discount rate” ρ (see Proposition 5) are:*

$$\zeta = 1 - \theta + \kappa^D, \quad \rho = \frac{\zeta}{\kappa}. \quad (52)$$

The price of equities is:

$$P_t = \frac{D_t}{\delta} e^{p_t}, \quad (53)$$

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:

$$p_t = b_f^p f_t, \quad b_f^p = \frac{1}{\zeta + \kappa\phi_f}. \quad (54)$$

Hence the variance of stock market returns is

$$\sigma_r^2 = \text{var} \left(\varepsilon_t^D + b_f^p \varepsilon_t^f \right), \quad (55)$$

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

$$\bar{\pi} = \gamma \sigma_r^2. \quad (56)$$

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

$$\pi_t = \bar{\pi} + b_f^\pi f_t, \quad b_f^\pi = -(\delta + \phi_f) b_f^p. \quad (57)$$

Finally, the interest rate is constant, and given by the consumption Euler equation (47):

$$r_f = -\ln \beta + \gamma g - \gamma(\gamma + 1) \frac{\sigma_y^2}{2}. \quad (58)$$

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, and that markets are inelastic. 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, and because the narrow framing makes the investor react to the variance of equity returns, rather than their covariance with consumption), as given in (56).

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 “flow-based” pricing kernel (see Section G.10 for details). In our general equilibrium model, this SDF is:

$$\mathcal{M}_{t+1} = \exp\left(-r_f - \pi_t \frac{\varepsilon_{t+1}^D}{\sigma_D^2} + \xi_t\right), \quad \pi_t = \bar{\pi} + b_f^\pi f_t, \quad (59)$$

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

This “flow-based” pricing kernel is an alternative to the consumption-based kernel of Lucas (1978). The core economics is in how flows affect prices, and the pricing kernel (59) just reflects that. The flow f_t modifies the price P_t according to Proposition 6 and also the pricing kernel \mathcal{M}_{t+1} , in such a way that $P_t = \mathbb{E}_t[\mathcal{M}_{t+1}(D_{t+1} + P_{t+1})]$ holds. The pricing kernel is in a sense a symptom rather than a cause in that market.

To sum up, the flow-based SDF (59) reacts to flows, and prices equities and bonds:

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

However, in this model, consumption does not directly price equities, though it does price bonds:

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

5.4 Calibration of the general equilibrium model

We now calibrate the general equilibrium model. This extends the calibration of Section 3.2, which is natural as the general equilibrium model is an extension of the basic infinite horizon model. We use the parameter values given in Table 5, which are all presented in annualized terms for clarity. We provide a summary discussion of our parameter choices here, leaving some details to Section H. Risk aversion is moderate, at $\gamma = 2$. The macroeconomic parameter values are standard, except for the pure rate of time preference.⁶⁰ 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 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.

Table 6 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 (with their 2.8% annual standard deviation) 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 7 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 featuring inelastic markets is competitive with other widely-used general equilibrium models that match equity market moments. Its main advantages, as we see it, are that it relies on an observable force, flows in and out of equities and that it

⁶⁰To get a small risk-free rate of 1% (and only for this reason), we need to make the agents very patient, so that $\beta > 1$. Indeed, this comes from the Ramsey equation (58), which is $r_f \simeq -\ln \beta + \gamma g$ (neglecting precautionary effects, which are very small in our calibration) with $\gamma g = 4\%$. We share this issue with the overwhelming majority of the macroeconomics literature: if we normalized the average growth rate to zero, like most of the macroeconomics literature, we would not have this difficulty. It would be easy to amend that, for example by adding a small probability of a disaster risk or by using Epstein-Zin preferences. We do not do that, because we do not wish to complicate the model.

Table 5: 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	$\theta = 87.5\%$
Mixed fund's sensitivity to risk premium	$\kappa = 1$
Speed of 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	$\gamma = 2$

Notes. Values are annualized.

Table 6: Moments generated by the calibration

Variable	Value
Macro elasticity	$\zeta = 0.16$
Macro elasticity with mean-reverting flow	$\zeta^M = \zeta + \kappa\phi_f = 0.2$
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	$\sigma_f = 2.8\%$
Slope of log price deviation to flow	$b_f^p = 5$
Slope of equity premium to flow	$b_f^\pi = -0.37$

Notes. Values are annualized.

Table 7: 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							
Horizon	Data			Model			
	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 $R_{t \rightarrow t+T} = \alpha_T + \beta_T \ln \frac{D_t}{P_t}$, 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.

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 highlighting how to think about inelastic market in general equilibrium.

6 Government Policy and Corporate Finance in Inelastic Markets

We now examine how a number of issues in finance change when markets are inelastic: government and corporate policies. Many readers may wish to skip to the conclusion, but in our experience a good fraction of readers will be interested in these topics.

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 $p = \frac{f^G}{\zeta}$.⁶¹ So,

⁶¹Note that we assume that investors do not change their holdings to counteract the government's holdings, meaning that Ricardian equivalence does not hold, perhaps because of a form of inattention to the government's actions (Gabaix (2020)).

if the government buys 1% of the market (which may represent roughly 1% of GDP), the market goes up by 5%.⁶²

This is what a number of central banks have done. In August 1998, the Hong Kong government, when it was under a 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 2020.⁶³ In inelastic markets, this may have a large price impact.⁶⁴ Those government purchases of equities offer a potentially attractive government policy, as they increase market values and lower the cost of capital for firms, and relax credit constraints. So, they might increase hiring and real investments by firms, and GDP. We think this is an interesting direction for future research.⁶⁵

6.2 Corporate finance in inelastic markets

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

In a frictionless model, this does not affect the firms’ values, as per Modigliani-Miller. In an inelastic model, it should now be clear that buybacks can 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 G.11). At time 0, we imagine the representative firm buys back a fraction b of the equity shares, where b is small (so that the new number of shares is $Q'_0 = (1 - b)Q_0$). The buyback is financed by a fall in the time-0 dividend, so the total dividend payout falls from \mathcal{D}_0 to $\mathcal{D}'_0 = \mathcal{D}_0 - P_0Q_0b$, where P_0 is the ex-dividend price, and P_0Q_0b is used to finance the share buyback.

We need to take a stance 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 F).

Proposition 7. (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*

⁶²If 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 (20). With the above calibration, this can be a moderate dampening if T is large enough.

⁶³Lockett, Hudson. “How the invisible hand of the state works in Chinese stocks.” *Financial Times*, 2/4/2020.

⁶⁴We are not aware of a quantification of the macro elasticity for Japan. Barbon and Gianinazzi (2019) and Charoenwong et al. (2020) quantify a micro elasticity – the differential impact on individual stocks that are owned versus not owned by the government.

⁶⁵Brunnermeier et al. (2020) caution about potentially adverse effect if the government’s purchases might become too central.

amount, hence keeping total payout constant at time 0. Then, the aggregate value of equities moves by a fraction

$$v = \frac{(\mu^D - \mu^G)\theta}{\zeta + \mu^G\theta}b, \quad (60)$$

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 provisional calibration Using the estimates of Di Maggio et al. (2020b), we set $\mu^D \simeq 0.5$ and $\mu^G \simeq 0.03$. Then, (60) says that a buyback of 1% of the market increases the market capitalization by 2.2%. The above papers (Baker et al. (2007); Di Maggio et al. (2020b)) do not exactly measure μ^D and μ^G : they measure 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 (respectively μ^G) is likely to be higher than the marginal propensity to consume out of dividends (respectively capital gains). In addition, what matters is the “long run” propensity, which is hard to measure, and one may conjecture that the long-run consumption adjustment to a lasting policy change will have $\mu^D - \mu^G$ closer to 0. One upshot is that it would be interesting for the empirical literature to estimate the long-run μ^D and μ^G , as it is important to understand the impact of firms’ actions such as buybacks in inelastic markets.

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 markets 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 because a permanent change in the demand for equities must permanently change their equilibrium prices – 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 enough elasticity to the market.” This is not true: in part because hedge funds are small (they own less than 5% of the market, see Section 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, those smart-money investors often face risk constraints and outflows 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 , which is a (signed) flow from bonds to equities.

“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 $q = -\zeta p + f$ has a component $-\zeta p$ expressing that “sellers” appetite to sell 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), as in (10).

“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 (20): 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.⁶⁶ 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

⁶⁶Indeed, $(1 + \rho)^{-T} = 99.8\%$, taking the ρ calibrated in Section 3.2 and $T = 1/52$ years.

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 can increase aggregate valuations (Section 6.2 detailed this). Hence, share buybacks are potentially a 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 informational channel. Most extant empirical evidence focuses on announcements at the firm level, while we emphasize their impact at the aggregate level. 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.

“Markets must be macro elastic as otherwise small flows would imply large price changes and market timing strategies would be too profitable.” The Sharpe ratios of market timing strategies depend on the properties of flows, see (23) and (24). If flows are highly persistent, prices may move a lot, but the per-period expected excess returns do not change much. Indeed, in the model in Section 5, the persistence of the dividend yield matches its empirical counterpart and using it for market-timing purposes does not work well out of sample.

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 (so that $\zeta > 0$, $\kappa = 0$), such as those of many funds that are 100% in equities and hence have zero elasticity, 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 may 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), ?). The pass-through between subjective beliefs and actions might be low, as it is for retail investors (Giglio et al. (2021a)). Finally, demand may respond little to prices because demand shocks are highly persistent.⁶⁷ In the

⁶⁷For instance, imagine a very simple model $f_t = \sum_k F_k \mathbb{I}(t \in [\tau_k^0, \tau_k^1])$, where F_k is constant, $[\tau_k^0, \tau_k^1]$ 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

end, while identifying the exact reasons for low market elasticity is 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 4.4, 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. (2021a).

To appreciate the richness of those determinants, let us observe that flow shocks could come from various 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 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 “delta hedging,” whereby traders who have sold put options and continuously hedge them need to sell stocks when stock prices fall.

Some further outstanding questions In addition to the two questions we just discussed, our framework makes a number of further issues interesting and researchable. For example, how much can and should governments intervene in equity markets? Do share buybacks account for a large share of market fluctuations? How forward-looking are the policies of funds (κ)? Generalizing, what are the cross-market elasticities, meaning the forces that create “contagion” across market? These same effects will also 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: Main proofs

Proof of Proposition 2 At time 0^- , before the inflow shocks, fund i ’s wealth is $\bar{W}_i = \bar{P}\bar{Q}_i + \bar{B}$, where $\bar{P}\bar{Q}_i$ and \bar{B}_i are respectively the fund’s holdings of equities and bonds:

$$\bar{P}\bar{Q}_i = \theta_i \bar{W}_i, \quad \bar{B}_i = (1 - \theta_i) \bar{W}_i.$$

At time 0, after the inflow shock, and the change in the equilibrium price to P , the fund’s wealth is $W_i = P\bar{Q}_i + \bar{B}_i + \Delta F_i$, so that $\Delta W_i = (\Delta P)\bar{Q}_i + \Delta F_i$. So, the value of the assets in the fund changes by a fraction:

$$w_i := \frac{\Delta W_i}{\bar{W}_i} = \frac{\bar{Q}_i \Delta P}{\bar{W}_i} + \frac{\Delta F_i}{\bar{W}_i} = \frac{\bar{P}\bar{Q}_i}{\bar{W}_i} \times \frac{\Delta P}{\bar{P}} + f_i = \theta_i \times p + f_i,$$

contrast with the traditional view in which flows have a temporary price impact (for instance Coval and Stafford (2007)).

that is:

$$w_i = \theta_i p + f_i. \quad (61)$$

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

Let us first take the case $\kappa_i = 0$. The demand (1) is:

$$Q_i = \frac{\theta_i W_i}{P} = \frac{\theta_i \bar{W}_i (1 + w_i)}{\bar{P} (1 + p)} = \bar{Q}_i \frac{1 + w_i}{1 + p},$$

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

$$q_i = \frac{Q_i}{\bar{Q}_i} - 1 = \frac{w_i - p}{1 + p} = \frac{\theta_i p + f_i - p}{1 + p} = \frac{f_i - \zeta_i p}{1 + p},$$

with $\zeta_i = 1 - \theta_i$. We see how $-\zeta_i$ is the (signed) demand elasticity, which includes crucial income effects.⁶⁸ For small price changes, this gives $q_i \simeq f_i - \zeta_i p$. We also see that, when $\kappa_i = 0$ for all funds, the equilibrium condition $q_S^D = 0$ leads to $p = \frac{f_S}{\zeta_S}$ exactly.

Next, consider the case with a general κ_i . 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 $\delta = \frac{D^e}{P} = 1 + r_f + \pi$, we have $\Delta \ln \frac{D^e}{P} = \frac{\Delta \pi}{1 + r_f + \pi} = \delta \hat{\pi}$ (with $\hat{\pi} = \Delta \pi$), so that:

$$\hat{\pi} = \delta (d - p). \quad (62)$$

We take logs in (1), so that $\ln Q_i = \ln W_i + \ln \theta_i - \ln P_i + \kappa_i \hat{\pi}$. Given that initially $\ln \bar{Q}_i = \ln \bar{W}_i + \ln \theta_i - \ln \bar{P}$, taking differences we have $\Delta \ln Q_i = \Delta \ln W_i - \Delta \ln P + \kappa_i \hat{\pi}$. Finally, we use the Taylor expansion $\Delta \ln W_i \simeq w_i$ and $\Delta \ln P \simeq p$ to yield:

$$q_i = w_i - p + \kappa_i \hat{\pi}. \quad (63)$$

Using (61), we obtain (7):

$$q_i = -(1 - \theta_i) p + f_i + \kappa_i \delta (d - p) = -(1 - \theta_i + \kappa_i \delta) p + f_i + \kappa_i \delta d.$$

Proof of Proposition 4 We call F_t the cumulative inflow into the mixed fund, normalizing F_0 to be the mixed fund's initial endowment of bonds. Then, as all dividend and bond coupon are given to the consumer, $W_t = P_t Q + F_t$, and in the baseline economy $\bar{W}_t = \bar{P}_t Q + \bar{F}_t$. We call $\tilde{F}_t := F_t - \bar{F}_t$ the deviation of the dollar flows from the baseline. Subtracting, we have $W_t - \bar{W}_t = (P_t - \bar{P}_t) Q + \tilde{F}_t$, i.e. $\bar{W}_t w_t = \bar{P}_t Q p_t + \tilde{F}_t$, so with $f_t = \frac{\tilde{F}_t}{\bar{W}_t}$,

$$w_t = \theta p_t + f_t. \quad (64)$$

⁶⁸This is the compensated or Hicksian elasticity of demand: indeed, after the price change, the fund can purchase its old holdings (which is the foundation of the Hicksian demand), simply because it already owns them). Controlling for fund wealth, the demand elasticity is -1 . But given fund wealth has an elasticity θ to the price, the total demand elasticity $(-\zeta)$ is $-1 + \theta$.

Now, from the demand for stocks, we have $Q_t P_t = W_t \theta e^{\kappa \hat{\pi}_t + \nu_t}$, while in the baseline economy $\bar{Q}_t \bar{P}_t = \bar{W}_t \theta$. Dividing through, we get: $\frac{Q_t P_t}{\bar{Q}_t \bar{P}_t} = \frac{W_t}{\bar{W}_t} e^{\kappa \hat{\pi}_t + \nu_t}$, so that $(1 + q_t)(1 + p_t) = (1 + w_t) e^{\kappa \hat{\pi}_t + \nu_t}$. Linearizing, $q_t + p_t = w_t + \kappa \hat{\pi}_t + \nu_t$. Hence, by (64),

$$q_t = -(1 - \theta)p_t + \kappa \hat{\pi}_t + f_t + \nu_t. \quad (65)$$

Finally, using $\hat{\pi}_t = \delta(d_t^e - p_t) + \mathbb{E}_t[\Delta p_{t+1}]$ (see (18)), we obtain $q_t = -(1 - \theta + \kappa \delta)p_t + \kappa \delta d_t^e + \kappa \mathbb{E}_t[\Delta p_{t+1}] + f_t + \nu_t$.

Proof of Proposition 5 Equation (19) can be rewritten as $q_t = \kappa(\mathbb{E}_t \Delta p_{t+1} - \rho p_t + \delta d_t^e) + f_t + \nu_t$. As $q_t = 0$, this is also:

$$\mathbb{E}_t \Delta p_{t+1} - \rho p_t + \delta d_t^e + \frac{f_t + \nu_t}{\kappa} = 0. \quad (66)$$

Defining $z_t := \delta d_t^e + \frac{f_t + \nu_t}{\kappa}$, 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 (65) with $q_t = 0$.

B Appendix: Identification methodology

We summarize the algorithms that we use to estimate the multipliers in Section 4.2 the multipliers and elasticities 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 investor-level data

We summarize the algorithm that we use to extract factors, η_t , in Section 4.2.

1. We run the panel regression

$$\Delta q_{it} = a_i + b_i \Delta y_t + c_t + \eta_{1t} x_{1i,t-1} + \eta_{2t} x_{2i,t-1} + \Delta \check{q}_{it},$$

where Δy_t is GDP growth, a_i is an investor fixed effect, c_t is a time fixed effect, $x_{1i,t-1}$ is lagged size, and $x_{2i,t-1}$ lagged active share. We collect the residuals, $\Delta \check{q}_{it}$.

2. We compute the time-series standard deviation of $\Delta \check{q}_{it}$ by investor. In each quarter, we sort investors into 20 groups based on this standard deviation. Intuitively, funds with different volatilities of $\Delta \check{q}_{it}$ are likely to have different exposures to the factors. By group and quarter, we average $\Delta \check{q}_{it}$, $\Delta \check{q}_{gt}^E$, where g indexes the groups.
3. We extract principal components based on the panel of 20 groups of $\Delta \check{q}_{gt}^E$.

B.2 Algorithm used for sector-level data

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

1. We construct pseudo-equal value weights $\tilde{E}_{i,t-1}$, where we start from $\tilde{E}_i^\sigma = \frac{\sigma_i^{-2}}{\sum_{k=1}^N \sigma_k^{-2}}$, where $\sigma_i = \sigma(\Delta q_{it})$, and define $\tilde{E}_i = \min \left\{ \xi \tilde{E}_i^\sigma, \frac{1.5}{N} \right\}$, where $\xi \geq 1$ is tuned so that $\sum_i \tilde{E}_i = 1$. We

exclude the corporate sector in constructing the instrument. This winsorizes the quasi-equal weights to be at most 50% higher than strict equal weights. This adjustment ensures that the equal weights are not too concentrated for sectors with very stable Δq_{it} .⁶⁹ This is relevant when the number of sectors is small, as is the case for the FoF.

2. We run the panel regression

$$\Delta q_{it} = \alpha_i + \beta_t + \gamma_i \Delta y_t + \delta_i t + \Delta \check{q}_{it}, \quad (67)$$

using \tilde{E} as regression weights, and construct the $\Delta \check{q}_{it}$ 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 in the subsequent period.

3. We extract the principal components of $\tilde{E}_i^{\frac{1}{2}} \Delta \check{q}_{it}$ and denote the estimated vector of principal components by $\eta_t^{PC,e}$.
4. We construct the GIV instrument:⁷⁰

$$Z_t = \sum_{i=1}^N S_{i,t-1} \Delta \check{q}_{it}. \quad (68)$$

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

$$\Delta p_t = \alpha + M Z_t + \lambda'_P \eta_t^e + e_t, \quad (69)$$

where $\eta_t^e = (\Delta y_t, \eta_t^{PC,e})$. 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

$$\Delta q_{Et} = \alpha_E - \zeta \Delta p_t + \lambda'_E \eta_t^e + e_t, \quad (70)$$

and the supply elasticity via

$$\Delta q_{Ct} = \alpha_C - \zeta_C \Delta p_t + \lambda'_C \eta_t^e + e_t. \quad (71)$$

References

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.

⁶⁹Quasi-equal weights \tilde{E}_j are preferable to equal weights $E_j = \frac{1}{N}$ as they add precision — in the same way in which to estimate a mean, weighing by inverse variance is better than equal weighing (Gabaix and Koijen (2020)). The primary objective of inverse variance weighing 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 as some sectors are very stable, the same concern applies, and we therefore winsorize the weights at $\frac{1.5}{N}$. While 50% is somewhat arbitrary, it is a significant departure from equal weights. We also explore the sensitivity of our results to this cutoff in Section D.4, and find them to be robust.

⁷⁰An equivalent way to proceed is to use $z_t = \sum_{i=1}^N S_{i,t-1} \check{u}_{it}$, where \check{u}_{it} 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 similar.

- Bacchetta, Philippe and Eric Van Wincoop**, “Infrequent portfolio decisions: A solution to the forward discount puzzle,” *American Economic Review*, 2010, *100* (3), 870–904.
- 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.
- Barberis, Nicholas, Ming Huang, and Richard H Thaler**, “Individual preferences, monetary gambles, and stock market participation: A case for narrow framing,” *American economic review*, 2006, *96* (4), 1069–1090.
- 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, Francesco Franzoni, and Rabih Moussawi**, “Do ETFs Increase Volatility?,” *The Journal of Finance*, 2018, *73* (6), 2471–2535.
- Ben-David, Itzhak, Francesco Franzoni, Rabih Moussawi, and John Sedunov**, “The granular nature of large institutional investors,” *Management Science*, forthcoming.
- Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song**, “Non-Fundamental Demand and Style Returns,” 2020. Working Paper.
- Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song**, “Style Investing, Positive Feedback Loops, and Asset Pricing Factors,” *Working Paper*, 2020.
- Ben-Rephael, Azi, Shmuel Kandel, and Avi Wohl**, “Measuring investor sentiment with mutual fund flows,” *Journal of Financial Economics*, 2012, *104* (2), 363–382. Special Issue on Investor Sentiment.
- 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, Julius Bonart, Jonathan Donier, and Martin Gould**, *Trades, quotes and prices: financial markets under the microscope*, Cambridge University Press, 2018.
- 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.
- Buffa, Andrea M, Dimitri Vayanos, and Paul Woolley**, “Asset management contracts and equilibrium prices,” Technical Report, National Bureau of Economic Research 2019.
- 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 H el ene 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.
- Cochrane, John H.**, “Presidential Address: Discount Rates,” *The Journal of Finance*, 2011, *66* (4), 1047–1108.
- Cole, Allison, Jonathan Parker, and Antoinette Schoar**, “Household Portfolios and Retirement Saving Over the Life Cycle,” 2021. Working Paper, MIT Sloan.
- 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.
- Di Maggio, Marco, Amir Kermani, and Kaveh Majlesi**, “Stock Market Returns and Consumption,” *Journal of Finance*, 2020, *75* (6), 3175–3219.
- Dierker, Martin, Jung-Wook Kim, Jason Lee, and Randall Morck**, “Investors’ Interacting Demand and Supply Curves for Common Stocks,” *Review of Finance*, 2016, *20* (4), 1517–1547.

- Dong, Xi, Namho Kang, and Joel Peress**, “How the Speed of Capital Affects Factor Momentum, Reversal and Volatility?,” *Available at SSRN*, 2021.
- Dou, Winston, Leonid Kogan, and Wei Wu**, “Common Fund Flows: Flow Hedging and Factor Pricing,” *Available at SSRN*, 2020.
- 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.
- Edelen, Roger M and Jerold B Warner**, “Aggregate price effects of institutional trading: a study of mutual fund flow and market returns,” *Journal of Financial Economics*, 2001, *59* (2), 195–220.
- 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,” *Working Paper*, 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,” Working Paper 28204, National Bureau of Economic Research December 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.
- Ghysels, Eric, Hanwei Liu, and Steve Raymond**, “Institutional Investors and Granularity in Equity Markets,” 2021. Working Paper.
- Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus**, “Five Facts About Beliefs and Portfolios,” *American Economic Review*, 2021, *111* (5), 1481–1522.

- Goetzmann, William N. and Massimo Massa**, “Index Funds and Stock Market Growth,” *The Journal of Business*, 2003, 76 (1), 1–28.
- 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.
- Haddad, Valentin, Paul Huebner, and Erik Loualiche**, “How Competitive is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing,” 2021. Working paper, UCLA.
- 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.
- Hartzmark, Samuel M. and David H. Solomon**, “Predictable Price Pressure,” 2022. Working paper, Chicago Booth.
- 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.
- Kahneman, Daniel and Dan Lovallo**, “Timid choices and bold forecasts: A cognitive perspective on risk taking,” *Management science*, 1993, 39 (1), 17–31.
- Kekre, Rohan and Moritz Lenel**, “Monetary policy, redistribution, and risk premia,” *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2020, (2020-02).
- 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 and Motohiro Yogo**, “Exchange Rates and Asset Prices in a Global Demand System,” 2020. NBER Working Paper 27342.
- 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.
- Li, Jiacui**, “Slow-moving liquidity provision and flow-driven common factors in stock returns,” *Available at SSRN 2909960*, 2018.
- Li, Jiacui**, “What Drives the Size and Value Factors?,” 2021. Working Paper, University of Utah.
- Li, Jiacui and Zihan Lin**, “Prices Are Less Elastic at More Aggregate Levels,” 2022. Working paper, David Eccles School of Business, University of Utah.
- Lou, Dong**, “A Flow-Based Explanation for Return Predictability,” *The Review of Financial Studies*, 12 2012, 25 (12), 3457–3489.
- 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.
- 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.
- Parker, Jonathan A, Antoinette Schoar, and Yang Sun**, “Retail Financial Innovation and Stock Market Dynamics: The Case of Target Date Funds,” *Working Paper*, 2020.
- Pavlova, Anna and Taisiya Sikorskaya**, “Benchmarking Intensity,” *Working Paper*, 2020.
- Peng, Cameron and Chen Wang**, “Factor demand and factor returns,” *Working Paper*, 2021.
- Petajisto, Antti**, “Why do demand curves for stocks slope down?,” *Journal of Financial and Quantitative Analysis*, 2009, 44 (5), 1013–1044.
- 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.
- Schmickler, Simon**, “Identifying the Price Impact of Fire Sales Using High-Frequency Surprise Mutual Fund Flows,” *Working Paper*, 2020.
- 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.
- 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

March 3, 2022

Table of contents

C	Data Sources and Construction	58
C.1	Sector-level data from the Flow of Funds	58
C.2	Investor-level data from 13F filings	62
D	Additional Empirical Results	65
D.1	Performance of the GIV Estimator: Simulations	65
D.2	Drawdown dynamics	66
D.3	Flows across investor classes are small	66
D.4	GIV estimates using FoF data: Robustness	68
D.5	The volatility of idiosyncratic demand shocks	70
D.6	Dynamics of mutual fund flows	70
D.7	The impact of mutual fund flows at longer horizons	70
D.8	Screening capital flows	71
E	Survey Details	74
F	Omitted Proofs	77
F.1	Derivation of (18)	77
F.2	Proof of Proposition 6	77
F.3	Proof of Proposition 7	77
F.4	Traditional rational or behavioral models predict that markets are extremely price-elastic	78
G	Theory Complements	82
G.1	The model with many asset classes	82
G.2	The model with time-varying market inelasticity	83
G.3	Micro versus macro elasticity: The cross-section of stocks	84
G.4	Short-term versus long-term elasticity when funds are inertial	86
G.5	On the link between the Kyle lambda and the market inelasticity	89
G.6	The model in continuous time	92
G.7	Infinite horizon model with heterogeneous funds	92
G.8	Complements to the general equilibrium model	93
G.9	Linking flows to beliefs	96

G.10 Pricing kernel consistent with flow-based pricing: Complements	100
G.11 Corporate finance in inelastic markets: Complements	103
G.12 When flows respond to risk premia in the long run	106
H Calibration of the general equilibrium model of Section 5: Details	112
H.1 From discrete to continuous time	112
H.2 Calibration steps	112
I Computing elasticities numerically in equilibrium models	114
I.1 Summary of the main results	114
I.2 General procedure	115
I.3 Lucas models with partially correlated labor income	116
I.4 Long run risk model	118

C Data Sources and Construction

C.1 Sector-level data from the Flow of Funds

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

C.1.1 Data items

We use Corporate Equities (Table 223) for equities. 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 equal to the total aggregate flow. The Flow of Funds revises historical data every quarter and we use the June 2019 vintage of the data.

C.1.2 Notation and data definitions

Sectors are indexed by $i = 1, \dots, I$, where $i = Foreign$ refers to the foreign sector. We observe holdings of equities, $W_{it}^{\mathcal{E}}$, 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_{it}^a , $a = W^{\mathcal{E}}, Tr, C, B$. Aggregate levels and flows omit the subscript i , implying, for instance, for equities $\sum_i W_{it}^{\mathcal{E}} = W_t^{\mathcal{E}}$ and for bonds $B_t = \sum_i B_{it}$. Lastly, the gross capital gain for equities is denoted by R_t^X and the return inclusive of dividend payments is denoted by R_t . We define the total assets of sector i as $W_{it} = W_{it}^{\mathcal{E}} + B_{it}$. Net issuances, $ni_t = \frac{NI_t}{W_{t-1}^{\mathcal{E}}}$, are based on equity markets.

In the FoF, equity flows are defined by $\Delta F_{it}^{\mathcal{E}} = W_{it}^{\mathcal{E}} - W_{i,t-1}^{\mathcal{E}} R_t^X$.⁷¹ We assume in what follows that the securities are adjusted at the end of the quarter, $\Delta F_{it}^a = \Delta F_{it}^a = (\Delta Q_{it}^a) P_t^a$, $a = \mathcal{E}, B$. The total per-period flow is $\Delta F_{it} = \Delta F_{it}^{\mathcal{E}} + \Delta F_{it}^B$ and in relative terms $\Delta f_{it}^{\mathcal{E}} = \frac{\Delta F_{it}^{\mathcal{E}}}{W_{i,t-1}^{\mathcal{E}}}$ for equities

⁷¹When 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.

and $\Delta f_{it}^B = \frac{\Delta F_{it}^B}{B_{i,t-1}}$ for bonds. The proportional per-period total flow is given by $\Delta f_{it} = \frac{\Delta F_{it}}{W_{i,t-1}}$. The equity shares are $S_{it} = \frac{W_{it}^\mathcal{E}}{W_t^\mathcal{E}}$. The relative change in equity demand, adjusted for price effects, is given by $\Delta q_{it}^\mathcal{E} = \Delta f_{it}^\mathcal{E} (R_t^X)^{-1} = \frac{\Delta Q_{it}^\mathcal{E}}{Q_{i,t-1}^\mathcal{E}}$. The aggregate per-period flow measure is defined as $\Delta f_{St} = \sum_i S_{i,t-1} \Delta f_{it}$.

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

C.1.3 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.⁷²

Let $W_{it}^{\mathcal{E},j}$ 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 $x_{US,t} := \sum_{i \in US} x_{it}$ for $x = W^\mathcal{E}, \Delta F^\mathcal{E}$, that is, for equity levels and equity flows.

We start from the following identities, for $j = D, F, T$,

$$x_{it}^D + x_{it}^F = x_{it}^T, \quad (72)$$

$$W_{it}^{\mathcal{E},j} = W_{i,t-1}^{\mathcal{E},j} R_t^{X,j} + \Delta F_{it}^{\mathcal{E},j}, \quad (73)$$

where $R_t^{X,j}$ is the capital gain as before. We observe x_t^D, x_t^F, x_t^T for $x = W^\mathcal{E}, \Delta F^\mathcal{E}$. 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:

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

It then follows that

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

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

For flows, we assume that

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

⁷²As US Treasuries are only issued in the US, obviously no adjustment is required for US Treasuries.

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

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

implying

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

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

C.1.4 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,” 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 sectors 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.

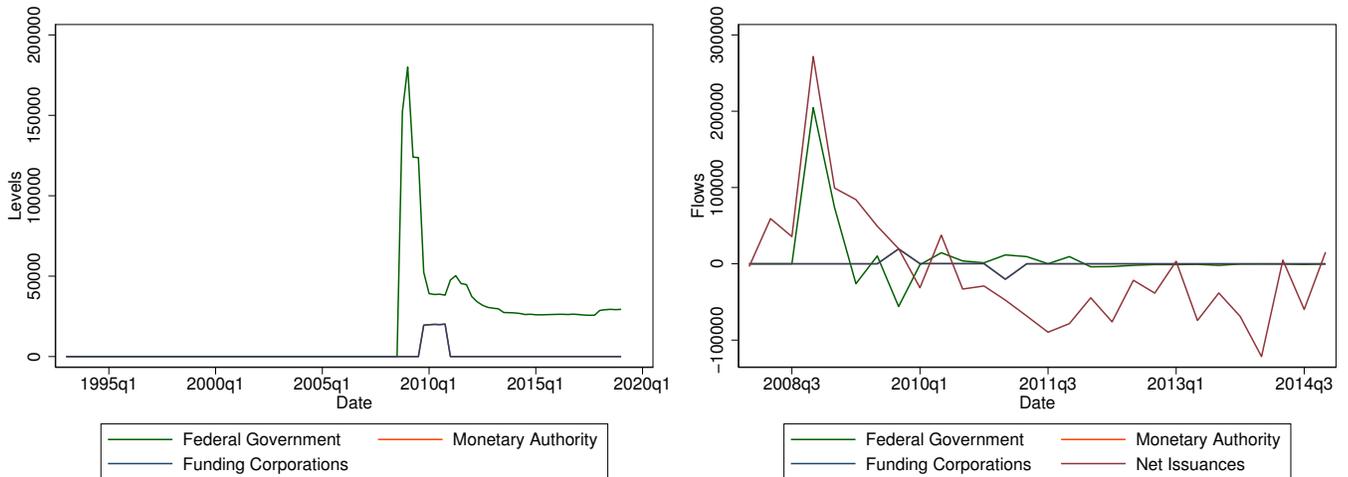
We aggregate the flows of these three sectors and subtract them from net issuances. We adjust the levels as well, and then remove these three sectors from our analysis.

C.1.5 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 zero flows since 1993 in most periods, 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.⁷³ These flows and holdings reflect firms’ holdings of other firms’ equity, for instance for strategic or speculative reasons. Prior to the September 2018

⁷³For details of the current procedure, please see here.

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. Levels and flows are in expressed in millions of nominal dollars.



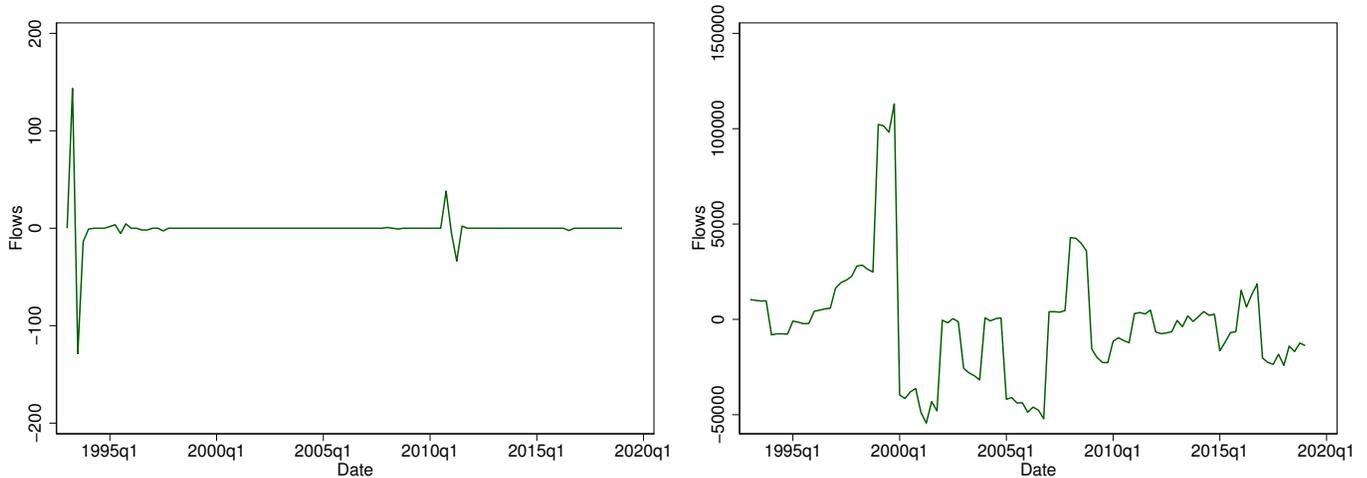
publication, the FoF showed the equity liability of the non-financial corporate sector net of these inter-corporate equity investments. The current release added the inter-corporate holdings as an asset and a liability. We undo this adjustment. For both sectors, we subtract the flows from net issuances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

C.1.6 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 holdings of 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 1993. In part, this starting date is driven by the fact that institutional ownership has been rising, and this allows us to disaggregate a large fraction of the holdings. Also, 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. 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.

Figure C.2: Flows of Foreign Banking Offices in the US and Non-financial Corporate Businesses. The left panel shows the flows for Foreign Banking Offices in the US and the right panel for Non-financial Corporate Businesses. The sample is from 1993.Q1 to 2018.Q4. Flows are in expressed in millions of nominal dollars.



C.1.7 Sample construction of the data for the GIV estimation

Before implementing the GIV procedure, we make two adjustments to the data 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 the data across time and sectors at the 5%- and 95%-percentiles for the period from 1993.Q1 to 2006.Q4 to mitigate the influence of outliers. This avoids the need to winsorize the data 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 as we show in Section D.4. We exclude the corporate sector in winsorizing the data.

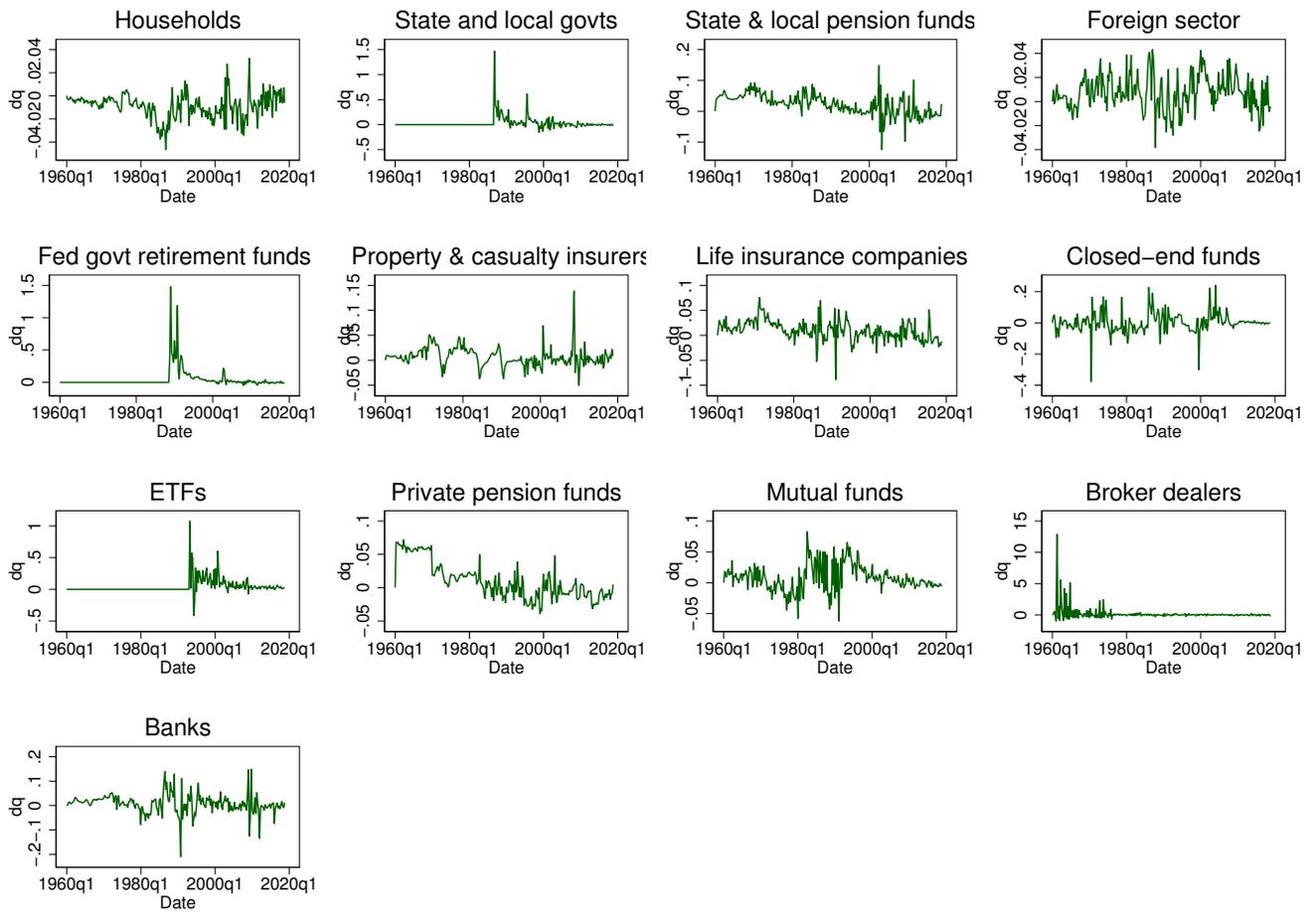
C.2 Investor-level data from 13F filings

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

C.2.1 Data construction

We source the 13F data from FactSet following Kojien 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`,

Figure C.3: Dynamics of equity flows across sectors. The figure shows the equity flows, $\Delta q_{it}^{\mathcal{E}}$, for the final 13 sectors in our sample from 1960.Q1 to 2018.Q4.



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.

C.2.2 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 change in equity demand as

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

where $\mathcal{E}_{it}^* = \frac{1}{2}(\mathcal{E}_{it}^- + \mathcal{E}_{i,t-1})$, 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, as in e.g. Davis and Haltiwanger (1992).

C.2.3 Constructing characteristics

Next, we discuss how we construct the characteristics that we use in Section B.1. We define the following characteristics

1. Log investor size, $\ln \mathcal{E}_{it}^*$.
2. Active share, which is defined as

$$\frac{1}{2} \sum_a |\theta_{iat} - \theta_{iat}^m|,$$

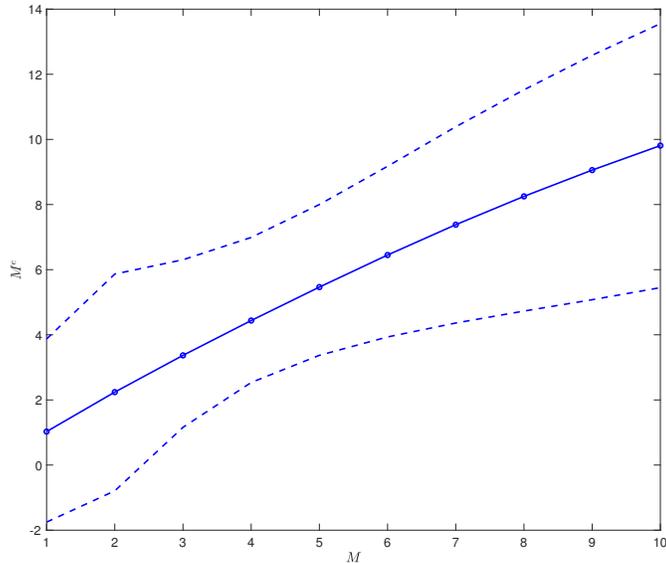
where θ_{iat} is the portfolio share and θ_{iat}^m the market-weighted portfolio of securities held by investor i at time t .

These characteristics define x_{it} , and we use their lagged values, $x_{i,t-1}$, to extract the factors.

C.2.4 Sample selection

We use 13F data to extract common factors, η_t , based on investors outside of the mutual fund industry using the same assignment of investor types as in Kojien et al. (2019). To mitigate the impact of outliers, we focus on the largest 1,000 investors in each period and we remove fund-quarter observations for which $|\Delta q_{it}| > 3\sigma_i$, where $\sigma_i = IQR_i/1.35$ and IQR_i is the interquartile range, a robust estimator of the standard deviation. We compute IQR_i using the full sample for a given investor. Lastly, we keep investors for which we have at least 20 observations after imposing these screens.

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.



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 one-factor factor model and estimate $f_t^\nu = \lambda \eta_t + u_t$ using principal components analysis. This provides us with an estimate of λ and $\sigma = \sigma(u_t)$. 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 ($\mu_\lambda = \iota' \lambda / N$), the average volatility of idiosyncratic shocks ($\mu_{\ln \sigma} = \iota' \ln \sigma / N$), and the dispersion in these parameters across investors ($\sigma_\lambda = \sigma(\lambda)$ and $\sigma_{\ln \sigma} = \sigma(\ln \sigma)$).

To simulate the data, we assume that $\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$, $\ln \sigma \sim N(\mu_{\ln \sigma}, \sigma_{\ln \sigma}^2)$, 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.

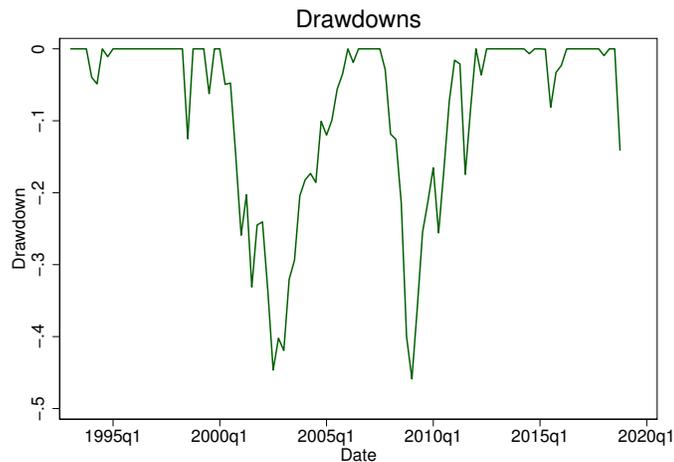
We consider 50,000 replications for each value of $\zeta = 0.1, 0.2, \dots, 1$ and report the average estimate alongside the 2.5% and 97.5% percentiles across all replications in Figure D.4. We report the multiplier M corresponding to the true data-generating process on the horizontal axis and the distribution 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.



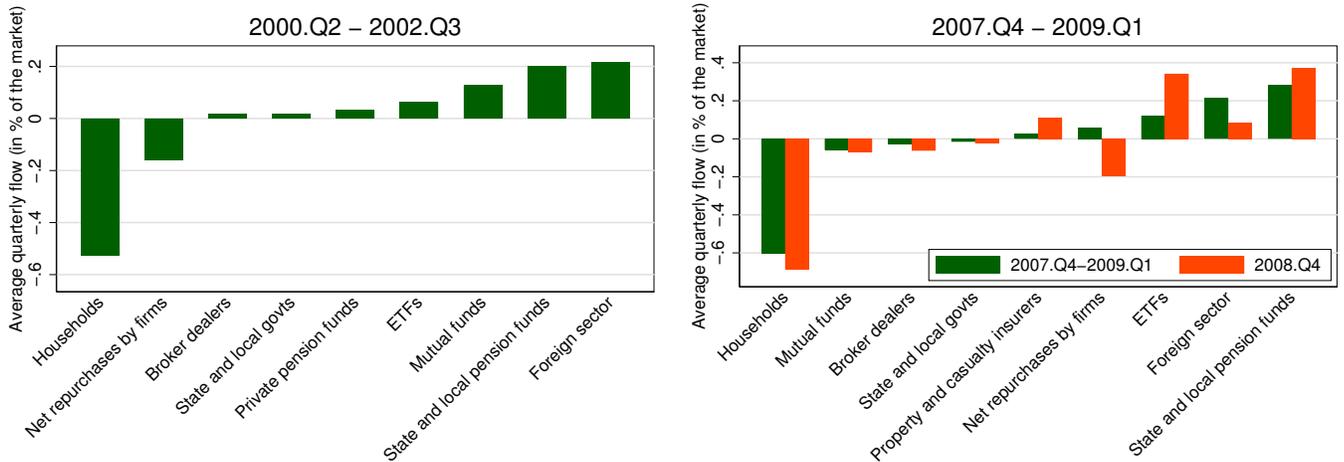
D.3 Flows across investor classes are small

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} , $\frac{\Delta F_{jt}^{\mathcal{E}}}{\mathcal{E}_{t-1}}$. We remove the mechanical effects due to revaluation, so that we show only the “active” flows. We then average the percent flow by sector across quarters for a given downturn.

The left panel of Figure D.6 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.⁷⁴

⁷⁴We 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.

Figure D.6: 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.



During the 2008 financial crisis, net repurchases by firms fell (as firms cut 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),⁷⁵ 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, this may reflect their mandate to maintain a fixed-share strategy instead of a conscious effort to time the market (see Proposition 2).⁷⁶

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 $y_t^{Gross} = \frac{\sum |\Delta F_{jt}^E| + |\Delta F_t^{Firm}|}{2\varepsilon_{t-1}}$, 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 $y_t^{AbsNet} = \frac{|\Delta F_t^{Firm}|}{\varepsilon_{t-1}}$, 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

⁷⁵During 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.

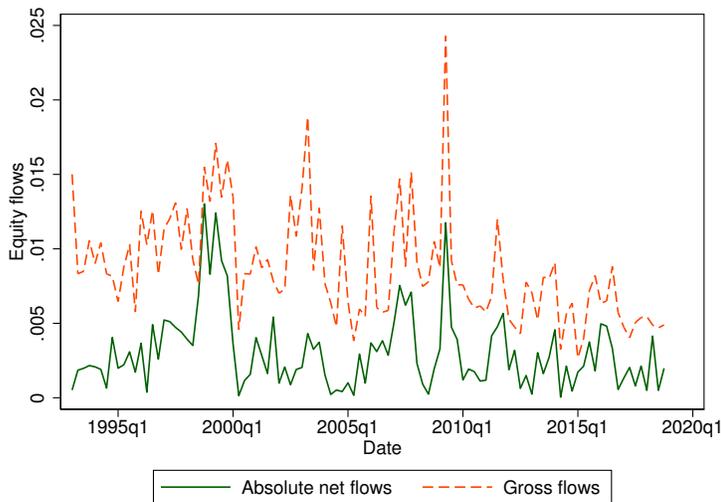
⁷⁶Relatedly, Timmer (2018) finds that in German data, banks (broker dealers) sell when stock prices fall, and pension funds buy.

markets across institutional sectors, which averages to approximately 0.6% per quarter.⁷⁷ We plot the time series of both measures in Figure D.7 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 N.3 also provides some additional first evidence on the link between flows and prices. Indeed, the demand by households (including mutual funds and 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.7: 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 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.

⁷⁷This 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.

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 (67). In column six, we control for the lagged value of Δq_{jt} in (67) and we allow for heterogeneous slope coefficients across sectors on $\Delta q_{j,t-1}$. 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 (67) and we allow for heterogeneous persistence coefficients across sectors on $\Delta q_{j,t-1}$. 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 (1.86)	8.00 (1.24)	6.94 (1.48)	7.65 (1.37)	6.63 (1.17)	6.77 (2.18)	6.79 (2.02)	7.17 (1.18)	5.12 (2.33)
GDP growth	5.99 (0.69)	5.99 (0.59)	6.06 (0.66)	6.02 (0.75)	6.14 (0.69)	6.01 (0.73)	6.20 (0.74)	5.99 (0.62)	5.97 (0.80)
η_1	21.06 (13.58)		22.09 (11.11)	15.59 (11.58)	32.54 (6.40)	25.24 (13.64)	-30.65 (15.65)	14.44 (10.71)	26.23 (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 3, 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 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.

Table D.9: Robustness of the GIV estimates for the Flow of Funds: Persistence in Z . The table replicates Table 3, 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 (1.96)	5.62 (1.21)				
Δp			-0.12 (0.03)	-0.16 (0.04)	-0.01 (0.01)	-0.01 (0.02)
Z (lag)	-1.59 (1.84)	-1.75 (1.46)	-0.30 (0.26)	-0.40 (0.29)	-0.02 (0.09)	-0.02 (0.08)
GDP growth	6.41 (0.93)	6.42 (0.90)	0.59 (0.25)	0.86 (0.24)	0.22 (0.13)	0.24 (0.12)
η_1	21.64 (13.74)	24.36 (13.10)	3.88 (1.58)	5.24 (2.42)	-0.72 (0.62)	-0.64 (0.66)
η_2		30.11 (6.02)		5.13 (1.16)		0.28 (0.80)
Constant	-0.01 (0.01)	-0.02 (0.01)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Observations	103	103	103	103	103	103

D.5 The volatility of idiosyncratic demand shocks

In Table D.11, we report the standard deviation of sector-specific demand shocks, u_{it} .

D.6 Dynamics of mutual fund flows

In Table D.12, we report the estimates of the dynamic model of mutual fund flows that we use in Section 4.2.

D.7 The impact of mutual fund flows at longer horizons

In Figure D.8, we repeat the analysis as in Section 4.3 but now using mutual fund flow innovations as in Section 4.2. We estimate the same regression as in (43), but now Z_t is constructed using data on mutual fund flows and η_t is estimated using 13F data. We present the results in Figure D.8. As before the impact of flow shocks on prices is persistent although the confidence interval is wide at longer horizons of one year.

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

D.8 Screening capital flows

In Table D.13, we report the slope estimates, β_j , of the regression in equation (44). 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.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 4.3 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: Dynamics of mutual fund flows. The table reports the dynamics of fund flows that we use in Section 4.2. We consider an AR(1), in Column 1, to AR(4) model of flows, in Column 4. In all cases we include a time trend. The standard errors, which correct for autocorrelation, are reported in parentheses.

	f	f	f	f
L.f	0.56 (0.04)	0.44 (0.04)	0.41 (0.04)	0.40 (0.04)
L2.f		0.22 (0.03)	0.18 (0.04)	0.17 (0.04)
L3.f			0.11 (0.04)	0.11 (0.04)
L4.f				0.03 (0.06)
t	-0.19 (0.03)	-0.15 (0.03)	-0.13 (0.02)	-0.12 (0.02)
Constant	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Observations	321	320	319	318
R^2	0.692	0.702	0.703	0.701

Figure D.8: 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 (43). 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 2000.Q1 to 2019.Q4.

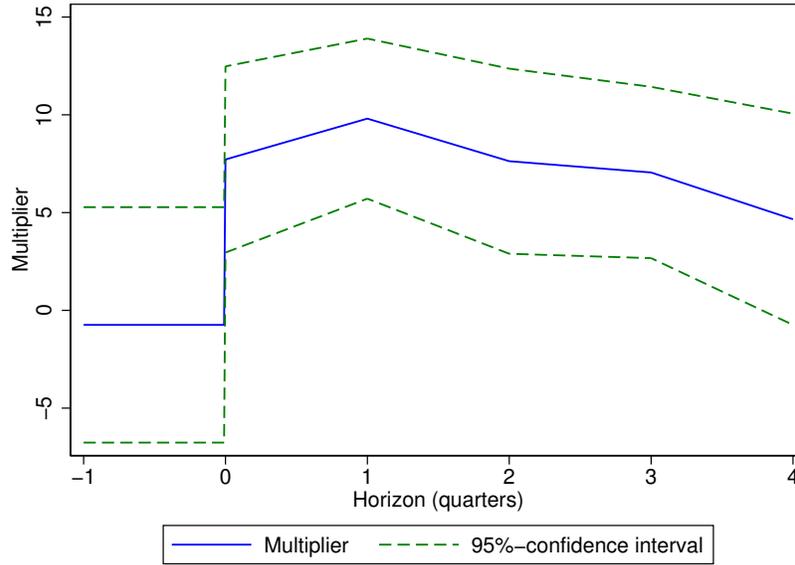


Table D.13: Assessing the mismeasurement of capital flows. The table reports the slope coefficient of the regression in equation (44) to assess whether capital flows are mismeasured. We consider a flow to be correctly measured when β_j is significantly different from one. The sample is from 1993.Q1 to 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

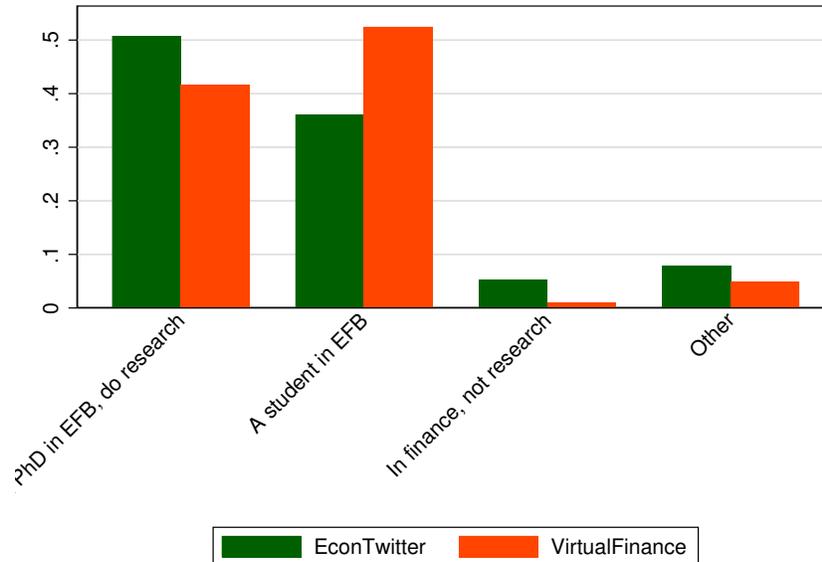
E Survey Details

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, 2020. We launched the Twitter survey on May 7, 2020. 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:⁷⁸
 - 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.

⁷⁸For 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 therefore inelastic demand.

Figure E.9: 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.org. In Figure E.9, 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 E.14, 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.⁷⁹ In Figure E.10, 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.

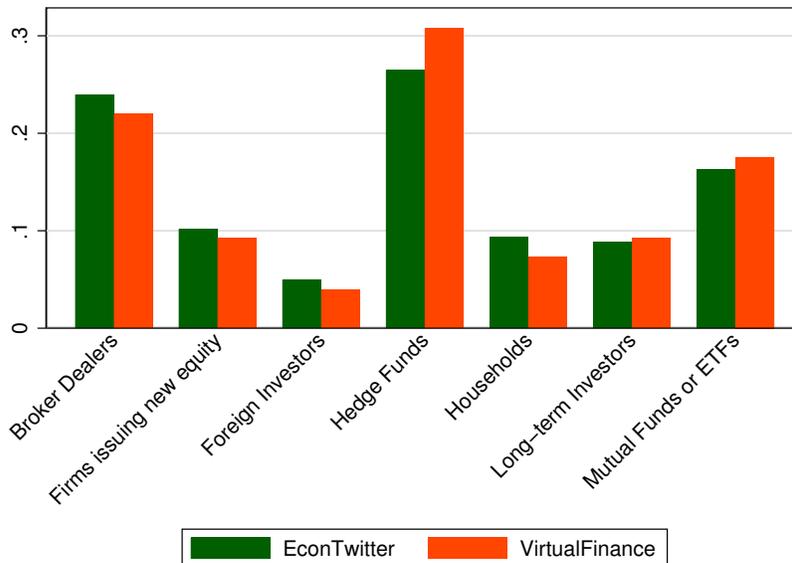
⁷⁹This approach to “testing the mechanism” is similar in spirit to the ideas in Chinco et al. (2020) .

Table E.14: 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.

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

Survey	Mean	Percentiles					
		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 E.10: 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.



F Omitted Proofs

F.1 Derivation of (18)

$$\begin{aligned}
1 + r_f + \bar{\pi} + \hat{\pi}_t &= 1 + r_f + \pi_t \\
&= \frac{\mathbb{E}_t [P_{t+1} + D_{t+1}]}{P_t} \\
&= \frac{\mathbb{E}_t [\bar{P}_{t+1} (1 + p_{t+1}) + \bar{D}_{t+1} (1 + d_{t+1})]}{\bar{P}_t (1 + p_t)} \\
&= \mathbb{E}_t \left[\frac{\bar{P}_{t+1}}{\bar{P}_t} (1 + p_{t+1} - p_t) + \frac{\bar{D}_{t+1}}{\bar{D}_t} \frac{\bar{D}_t}{\bar{P}_t} (1 + d_{t+1} - p_t) \right] \\
&= \mathbb{E}_t [(1 + g) (1 + p_{t+1} - p_t) + (1 + g) \delta (1 + d_{t+1} - p_t)] \tag{74} \\
&= (1 + g) (1 + \delta) + (1 + g) \mathbb{E}_t [\Delta p_{t+1} + \delta (d_{t+1} - p_t)] \tag{75}
\end{aligned}$$

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

$$\hat{\pi}_t = (1 + g) \mathbb{E}_t [\Delta p_{t+1} + \delta (d_{t+1} - p_t)] \tag{76}$$

In the text, to reduce the notational clutter, we take (18), which is this expression using the definition of δ as the baseline (i.e., frictionless) value of $\frac{\mathbb{E}_t [D_{t+1}]}{P_t}$ rather than of $\frac{D_t}{P_t}$. This can be also 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 $\kappa (1 + g)$.

F.2 Proof of Proposition 6

The values of p_t , $\hat{\pi}_t$ and so on, were derived in (26). We need to derive the flow f_t , the interest rate r_f , and the average risk premium $\bar{\pi}$.

As $\bar{W}_t = \frac{1}{\theta} \frac{D_t}{\delta}$ the baseline value of the mixed fund (when the behavioral friction is 0), (17) and (49) give: $f_t = \frac{F_t - \bar{F}_t}{\bar{W}_t} = \frac{b_t \frac{D_t}{\delta}}{\frac{1}{\theta \delta} D_t} = \theta b_t$.

Next, 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, $\mathbb{E}_t [V^p (R_{t+1})]$, with $V^p (R) = \frac{R^{1-\gamma}-1}{1-\gamma}$. Then, approximately, the allocation in equities is $\bar{\theta}^E = \frac{\bar{\pi}}{\gamma \sigma_r^2}$. Given that in equilibrium all the wealth comes from equity, the equity premium is $\bar{\pi} = \gamma \sigma_r^2$.

F.3 Proof of Proposition 7

Calling \mathcal{D}_1 the aggregate dividend, the dividend per share goes from $D_1 = \frac{D_1}{Q_0}$ to $D_1' = \frac{D_1}{Q_0'} = \frac{D_1}{1-b}$. So, the time-1 dividend per share increases by a fraction $d = b$.

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 $v = \Delta \ln (PQ) = p + q^S$ the

change in the market value of the firm, and r the excess return created by the buyback, we have:

$$\text{Frictionless model: } q^S = -b, \quad d = b, \quad p = b, \quad v = 0, \quad r = 0. \quad (77)$$

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 \mathcal{D}_0 to $\mathcal{D}_0 - P_0 Q_0 b$. So the households experience a change in dividend received $\Delta \mathcal{D}_0 = -P_0 Q_0 b$ (recall that the all dividends are passed on by the fund to the consumers)⁸⁰ and a capital gain $Q_0 \Delta P = P_0 Q_0 p$. Recall that we said that 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. This means that the flow is

$$\Delta F_0 = (1 - \mu^D) \Delta \mathcal{D}_0 - \mu^G Q_0 \Delta P \quad (78)$$

Indeed, if the extra dividend received is $\Delta \mathcal{D}_0$, there is a counteracting flow of $(1 - \mu^D) \Delta \mathcal{D}_0$, so that the total dividend change removed from the mixed fund is $\mu^D \Delta \mathcal{D}_0$. Likewise $\mu^G Q_0 \Delta P$ is “removed” from the mixed fund. This means that the flow is

$$f = \frac{\Delta F_0}{W_0} = \frac{Q_0 P_0}{W_0} \frac{(1 - \mu^D) \Delta \mathcal{D}_0 - \mu^G Q_0 \Delta P}{Q_0 P_0} = \theta [(1 - \mu^D) (-b) - \mu^G p]$$

The total demand change by the mixed fund is then $q = -\zeta p + \kappa \delta d + f$, and should be equal to the supply change $q^S = -b$. So

$$0 = q - q^S = -\zeta p + \kappa \delta d + f + b = -\zeta p + \kappa \delta b - \theta (1 - \mu^D) b - \theta \mu^G p + b$$

and the share price change is:

$$p = \frac{\zeta + \mu^D \theta}{\zeta + \mu^G \theta} b. \quad (79)$$

This yields (60) and implies that $p > b$ if $\mu^D > \mu^G$. A share buyback increases the market value by $v = p + q^S = p - b > 0$.

F.4 Traditional rational or behavioral models predict that markets are 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

⁸⁰The outcome would be the same if the rule was different. The total dividend change removed from the mixed fund would still be $\mu^D \Delta \mathcal{D}_0$.

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.⁸¹ How much capital should flow into equities? The next proposition answers this (the proof is at the end of this subsection).

Proposition 8. (Market elasticity in frictionless rational or behavioral models) *We derive the 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:*

$$\zeta^r = \frac{1}{\pi} \frac{C}{W^\varepsilon}, \quad (80)$$

where π is the equity premium, C is aggregate consumption and W^ε is the stock market capitalization. In a disaster model where growth rates follow a jump process, the elasticity of demand for equities is $\zeta^{r,D} = \frac{1}{\pi} \frac{C}{W^\varepsilon} \frac{(1-B^\gamma)B}{\gamma(1-B)}$, where B is the recovery rate of the endowment after a disaster.

Take the calibrated values $C = 0.8Y$, where Y is GDP, $W^\varepsilon = Y$ (as the typical market capitalization is roughly equal to GDP), and $\pi = 4\%$. Then (80) 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:

$$\frac{\zeta^r}{\zeta} = 100. \quad (81)$$

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 (80) 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 8 is $\zeta^{r,D} = 8$, so

⁸¹Johnson (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 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 8), 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.

that it is 40 times larger than the empirical one. In Section I we verify numerically that a similar reasoning works for long run risks models (Bansal and Yaron (2004)). We suspect that it would also apply to a habit formation model (Campbell and Cochrane (1999)).^{82,83}

Proof of Proposition 8 *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.

$$d\theta = \frac{d\pi}{\pi}. \quad (82)$$

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:

$$\frac{\Delta Q}{Q^\mathcal{E}} = \frac{\Delta\pi}{\psi\pi}. \quad (83)$$

If the value of equity changes by p , the equity premium changes by $\Delta\pi = -\delta p$ (see (18)), 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 $D_t Q^\mathcal{E} = \psi Y_t$.⁸⁴ So,

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

⁸²One could imagine other models, with idiosyncratic risk, but that would take us far afield.

⁸³As a correlate, traditional models counterfactually predict very correlated flows and beliefs. Indeed, with several institutions and a demand $q_{it} = -\zeta^r p_t + f_{it}^\nu$, 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_{it}^ν 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.

⁸⁴This is true in the logic of the Lucas tree model with no investment. In the calibration, we take the ‘‘fruit’’ to be consumption, which less than GDP as there is investment (in a closed economy), and indeed we take $C_t = 0.8Y_t$.

which is the announced expression.

Case 2: Disaster risk. The reasoning is the same, except that expression (82) 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:

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

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

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

so the FOC characterizing the equity share is:

$$\mathbb{E} \left[(1 + \theta \tilde{r}_{t+1})^{-\gamma} \tilde{r}_{t+1} \right] = 0. \quad (84)$$

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 (84) becomes

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

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

$$\pi = p^D (1 - \theta (1 - B))^{-\gamma} (1 - B). \quad (85)$$

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

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

i.e.

$$d\theta = \frac{d\pi}{\pi} \frac{B}{\gamma (1 - B)} \quad (86)$$

Finally, as $\pi = p^D B^{-\gamma} (1 - B)$ by (85), the risk premium in a full sample (including an average number of disasters) is: $\bar{\pi} = \pi - p^D (1 - B) = p^D (B^{-\gamma} - 1) (1 - B)$, so $\frac{\bar{\pi}}{\pi} = 1 - B^\gamma$. So (86) gives

$$d\theta = \frac{d\pi}{\bar{\pi}} \frac{B (1 - B^\gamma)}{\gamma (1 - B)}, \quad (87)$$

which is the disaster counterpart to (82): 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 (82) by (87).

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).

G Theory Complements

G.1 The model with many asset classes

We can easily extend the model to K asset classes, indexed by $A \in \{1, \dots, K\}$, 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.

Two-period model 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):

$$P_A Q_A^D = \theta_A W \exp \left(\sum_{B=1}^K \kappa_{AB}^D (d_B - p_B) \right).$$

For instance, if $\kappa_{AB}^D = 0$ the mixed fund seeks to keep a constant share θ_A in asset A . When κ_{AB}^D 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 class A changes by a fraction⁸⁵

$$q_A^D = - \sum_B \zeta_{AB} p_B + f_A + \sum_B \kappa_{AB}^D d_B, \quad (88)$$

where the cross-elasticities of demand ζ_{AB} express how demand for asset A changes with a change in the price of asset B : $\zeta_{AB} = 1_{A=B} - \theta_B + \kappa_{AB}^D$. In vector form, this gives

$$q = -\zeta p + f + \kappa^D d, \quad (89)$$

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):

$$p = \zeta^{-1} (f + \kappa^D d). \quad (90)$$

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.

Infinite horizon model The formulas of the papers extend again, replacing scalars by matrices. For instance, the demand is

$$q_t = -\zeta p_t + f_t + \kappa^D d_t^e + \kappa \mathbb{E}_t [p_{t+1} - p_t],$$

⁸⁵With just one fund, f_A is the same across assets classes A . But with several funds, by aggregation, f_A differs across asset classes so we use that more general notation.

where q_t, f_t, d_t, p_t are vectors and κ, κ^D, ζ are matrices. In equilibrium, $q_t = 0$ at all dates, so that

$$p_t = (\zeta + \kappa)^{-1} (\kappa \mathbb{E}_t [p_{t+1}] + f_t + \kappa^D d_t^e).$$

Defining the matrix ρ :

$$\rho := \kappa^{-1} \zeta, \tag{91}$$

we have

$$p_t = a (f_t + \kappa^D d_t^e) + (1 + \rho)^{-1} \mathbb{E}_t [p_{t+1}]$$

with $a = (\zeta + \kappa)^{-1}$. Solving forward, we have the multi-asset equivalent of (20):

$$p_t = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{1}{(1 + \rho)^{\tau-t}} (a f_\tau + a \kappa^D d_\tau^e). \tag{92}$$

In the AR(1) case where $f_t = (1 - \phi_f) f_{t-1} + \varepsilon_t^f$, with ϕ_f a matrix now, we have $p_t = b_f^p f_t$ with b_f^p a matrix equal to:

$$b_f^p = \sum_{h=0}^{\infty} (1 + \rho)^{-h} a (1 - \phi_f)^h. \tag{93}$$

In general, there is no closed form, but b_f^p can be computed iteratively as:

$$b_f^p = a + (1 + \rho)^{-1} b_f^p (1 - \phi_f). \tag{94}$$

G.2 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 $\mathbb{E}_t d_\tau^e = 0$. Then, we have the following variant of Proposition 5 (the derivation is similar):

$$p_t = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho}{(1 + \rho)^{\tau-t+1}} \frac{f_\tau + \nu_\tau}{\zeta_\tau}. \tag{95}$$

To be concrete, we study the case

$$\frac{1}{\zeta_t} = \frac{1}{\zeta} (1 + \mathcal{M}_t), \quad \mathbb{E}_t \mathcal{M}_{t+1} = (1 - \phi_\zeta) \mathcal{M}_t,$$

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, at $t \geq 0$,

$$p_t = \frac{f_0}{\zeta} \left(1 + \frac{\rho}{\rho + \phi_\zeta} \mathcal{M}_t \right). \tag{96}$$

So, if the flow f_0 happens during a time of high market inelasticity $\frac{1}{\zeta_0}$ (i.e. high \mathcal{M}_0), then the price impact is higher, which makes sense. It is the average future value of the inelasticity shifter ($\frac{\rho}{\rho + \phi_\zeta} \mathcal{M}_0$) that matters, rather than the current inelasticity shifter (\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 that returns that happened during a high-volatility period mean-revert faster.

A tentative calibration. With $\rho = 0.16/\text{year}$ and $\phi_\zeta = 0.15/\text{year}$, we have $\frac{\rho}{\rho + \phi_\zeta} \simeq 0.5$, so have then to get a price impact higher by a factor 0.5, we need $\mathcal{M}_t = \frac{0.5}{0.5} = 1$, i.e. a halving of ζ_t . This effect might be detectable, though not easily.

G.3 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 ζ^\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.

G.3.1 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 *asset*), which comes from a “tracking error” type of mandate: the fraction in equities allocated to asset a is

$$\frac{P_{at}Q_{at}}{P_tQ_t} = \theta_a^\mathcal{E} e^{\kappa^\perp \hat{\pi}_{at}^\perp + \theta^\perp p_a^\perp + \nu_a^\perp}. \quad (97)$$

Indeed, $P_{at}Q_{at}$ is the dollar demand for asset a , and P_tQ_t is the dollar demand for the aggregate stock market. On average, their ratio is $\theta_a^\mathcal{E}$. The term κ^\perp is the micro-elasticity of demand with respect to the anomalous part of the equity premium $\hat{\pi}_{at}^\perp$. The term θ^\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 $q_{at} = \frac{Q_{at}}{Q_a} - 1$ the deviation of the demand from the baseline, and $q_{at}^\perp = q_{at} - q_t$ how much asset a deviates from the baseline, we obtain the following counterpart to Proposition 4 (the proof is in Appendix F).

Proposition 9. (*Demand for individual stocks in the infinite-horizon model*) *The demand change (compared to the baseline) for an individual asset a is $q_{at} = q_t + q_{at}^\perp$, where q_t is the demand change for the aggregate stock market seen in Proposition 4, and q_{at}^\perp is the asset- a specific demand change, given by*

$$q_{at}^\perp = -\zeta^\perp p_{at}^\perp + \kappa^\perp \delta d_{at}^{e,\perp} + \kappa^\perp \mathbb{E}_t [\Delta p_{a,t+1}^\perp] + \nu_{at}^\perp \quad (98)$$

where ζ^\perp is the micro-elasticity of demand for individual stocks:

$$\zeta^\perp = 1 - \theta^\perp + \kappa^\perp \delta. \quad (99)$$

Proof. Equation (97) implies

$$q_{at}^\perp = -(1 - \theta^\perp) p_{at}^\perp + \kappa^\perp \hat{\pi}_{at}^\perp + \nu_{at}^\perp. \quad (100)$$

Likewise, the analogue of (18) is $\hat{\pi}_{at}^\perp = \mathbb{E}_t [\Delta p_{a,t+1}^\perp] + \delta (d_{at}^{e,\perp} - p_a^\perp)$, with $d_{at}^{e,\perp} := \mathbb{E}_t [d_{a,t+1}^\perp]$. Combining the two gives the announced expression. \square

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 by ζ^\perp , p_{at}^\perp , q_{at}^\perp , and so on. For

instance, the equilibrium value of the price p_{at}^\perp is as in Proposition 5, replacing ν_t by ν_a^\perp . We next spell this out and draw consequences. See Betermier et al. (2019) for an alternative demand-based model of the cross-section of stocks. Micro-elasticity of demand versus macro-elasticity of demand

Suppose that there is a “stock specific flow”, whereby someone buys ΔF_a^\perp worth of stock a , while selling ΔF_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 $q_{at}^\perp + f_a^\perp = 0$. So, the impact of a flow is:

$$p_a^\perp = \frac{f_a^\perp}{\zeta^\perp}, \quad (101)$$

where ζ^\perp is the price micro-elasticity of demand (99). 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, ζ^\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 $\kappa^\perp = 5$.⁸⁶

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

$$p_a = \frac{f_a}{\zeta^\perp} + \left(\frac{1}{\zeta} - \frac{1}{\zeta^\perp} \right) \omega_a f_a. \quad (102)$$

⁸⁶This ratio of price impact of roughly 1 to 5 is also consistent with Benzaquen et al. (2017).

For the other stocks $b \neq a$, we have $f_b^\perp = -f = -\omega_a f_a$, so the impact is:

$$p_b = \left(\frac{1}{\zeta} - \frac{1}{\zeta^\perp} \right) \omega_a f_a, \quad \text{for a stock } b \neq a. \quad (103)$$

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 \$1.2,⁸⁷ 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 (20) in asset-specific space:

$$p_{a,t}^\perp = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho^\perp}{(1 + \rho^\perp)^{\tau-t+1}} \left(\frac{f_{a\tau}^\perp}{\zeta^\perp} + M^{D,\perp} d_{a\tau}^{e\perp} \right). \quad (104)$$

Conclusion: Aggregate versus cross-section We conclude that the aggregate model extends 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 ζ^\perp , the main object of interest in this study is the aggregate elasticity ζ .

G.4 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” demand q_t^a :

$$q_t = q_t^a + f_t.$$

We model the actual active demand with inertia as:

$$\Delta q_t^a = \mu \Delta q_t^{a,\nu} + \phi \Delta t (q_{t-1}^{a,\nu} - q_{t-1}^a), \quad (105)$$

where $q_t^{a,\nu}$ is the “virtual active demand” – the one of a non-inertial fund:

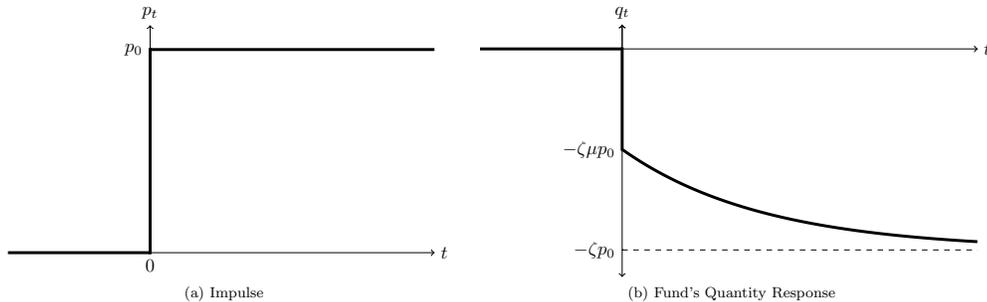
$$q_t^{a,\nu} = -(1 - \theta) p_t + \kappa \hat{\pi}_t + \nu_t = -\zeta p_t + \kappa (\mathbb{E} p_{t+1} - p_t) + \kappa \delta d_t + \nu_t, \quad (106)$$

with $\mu \in [0, 1]$ and $\phi \geq 0$. A frictionless investor has $\mu = 1$. The lower μ and ϕ , the more frictional the investor. 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).

⁸⁷Indeed, $\frac{1}{\zeta^\perp} + \left(\frac{1}{\zeta} - \frac{1}{\zeta^\perp} \right) \times \omega_a = 1 + (5 - 1) \times 5\% = 1.2$.

Figure G.11: This figure shows the quantity adjustment of an inert fund after a change in the aggregate stock price. It illustrates Proposition 10. 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.



Proposition 10. (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:*

$$q_t = -\zeta (1 + (\mu - 1)(1 - \phi)^t) p_0, \quad (107)$$

while its virtual demand is $q_t^\nu = -\zeta p_0$.

Figure G.11 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. We also assume $\phi < \mu$, which is automatically true in the limit of small time intervals.⁸⁸

Proposition 11. (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 $t \geq 0$),*

$$p_t = M_t f_0, \quad M_t = \frac{1}{\zeta} + b(1 - \Phi)^t \quad (108)$$

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 (108). We normalize $f_0 = -1$. Plugging this in (106) gives

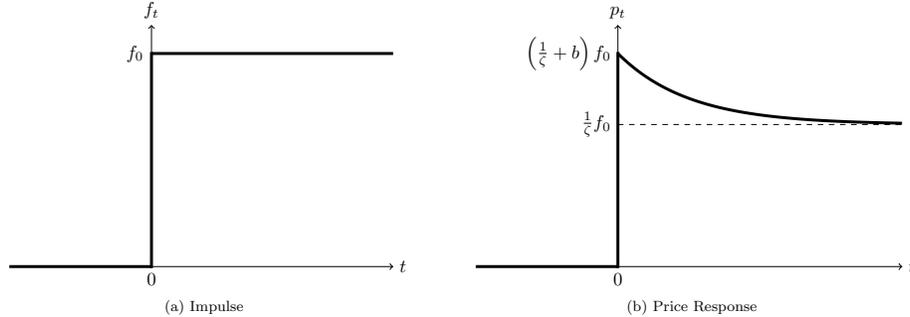
$$q_t^{a,v} = (1 + b\zeta(1 - \Phi)^t) + b\kappa\Phi(1 - \Phi)^t = 1 + b(\zeta + \kappa\Phi)(1 - \Phi)^t = 1 + c(1 - \Phi)^t.$$

For $t \geq 0$, equilibrium imposes $q_t^a + f_0 = 0$, i.e. $q_t^a = 1$. So (105) gives, for $t > 0$

$$0 = \Delta q_t^a = \mu \Delta q_t^{a,\nu} + \phi (q_{t-1}^{a,v} - q_{t-1}) = c(-\Phi\mu + \phi)(1 - \Phi)^{t-1},$$

⁸⁸In the limit of small time intervals, we replace $(1 - \Phi)^t$ in (108) by $e^{-\Phi t}$.

Figure G.12: This figure shows the price dynamics caused by an unanticipated time-0 demand shock, when investors are inertial. It illustrates Proposition 11. 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 investors are not inert, $b = 0$ and the price immediately jumps to its permanent value p_∞ .



which leads to $\Phi = \frac{\phi}{\mu}$. At time 0, (105) gives

$$q_0^a = \mu q_0^{a,v} = \mu (1 + b (\zeta + \kappa \Phi)).$$

As $q_0^a = 1$, this gives $b = \frac{1-\mu}{\mu(\zeta + \kappa \Phi)}$. □

Figure G.12 illustrates the dynamics of (108). An unanticipated, permanent inflow f_0 at time 0 has an immediate price impact $(\frac{1}{\zeta} + b) f_0$ that is bigger than the long-run price impact $\frac{f_0}{\zeta}$. The initial “excess reaction” $b f_0$ dies down at the exponential rate Φ . When funds are not inert, $b = 0$. This echoes the findings in Duffie (2010), with a somewhat different model.

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 .⁸⁹

Calibration We discuss the model calibration.⁹⁰ 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 $\zeta = 0.16$ (with $\zeta^M = 0.2$) and $\kappa = 1$ year for illustration:

$$b = \frac{1 - \mu}{\mu \zeta + \phi \kappa} = \frac{1 - 0.5}{0.5 \cdot 0.16 + 1 \cdot 1} \simeq 0.45.$$

The immediate price impact is $\frac{1}{\zeta} + b = 5.4$, 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 $b f_t (1 - \Phi)^t$,

⁸⁹Imagine 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 .

⁹⁰We use a “continuous time” calibration: The expressions work in continuous time, which makes calibration easier, replacing expressions like $(1 - \phi)^t$ by $e^{-\phi t}$.

the premium is $b\Phi f = 0.45 \cdot 2 \cdot 0.5\% = 0.4\%$ (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 (20) and the price impact with inertia (108).

Proposition 12. (Price impact with inertial funds) *When funds exhibit inertia, the price impact of inflows df_s is:*

$$p_t = \frac{f_{-\infty}}{\zeta} + \sum_{s=-\infty}^{\infty} G(t-s) \mathbb{E}_t[\Delta f_s], \quad (109)$$

where

$$G(\tau) = \begin{cases} \left(\frac{1}{\zeta} + b\right) \frac{1}{(1+\rho)^\tau} & \text{if } \tau < 0, \\ \frac{1}{\zeta} + b(1-\Phi)^\tau & \text{if } \tau \geq 0. \end{cases} \quad (110)$$

with $\rho = \zeta/\kappa$ is in (21), and b is in Proposition 11. When there is no inertia, $b = 0$.

Proof. This can be checked by the “plug and verify” method for market clearing, $q_t = 0$. □

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 (108) holds, with more complex expressions for b and Φ .

G.5 On the link between the Kyle lambda and the market inelasticity

G.5.1 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 micro structure regression is, as in Evans and Lyons (2002); Hasbrouck (2007):

$$p_t - p_{t-1} = \lambda(\Delta f_t - \mathbb{E}_{t-1}[\Delta f_t]) \quad (111)$$

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, we suppose some autocorrelation in the order flow (like Madhavan et al. (1997), Lillo et al. (2005) and Bouchaud et al. (2018)):

$$\Delta f_t = (1 - \phi_g) \Delta f_{t-1} + \varepsilon_t, \quad (112)$$

where ε_t is i.i.d. So, an innovation ε_t creates an innovation to the eventual cumulative flow:⁹¹

$$\lim_{h \rightarrow \infty} \mathbb{E}_t [f_{t+h} - f_{t-1}] = K \varepsilon_t, \quad K = \frac{1}{\phi_g}.$$

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” meta-orders by two other funds, then $K = 3$.⁹² The two forces combine: if a fund splits its meta-order 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,⁹³

$$\Delta p_t = \frac{K}{\zeta} \varepsilon_t = \frac{K}{\zeta} (\Delta f_t - \mathbb{E}_{t-1} [\Delta f_t]).$$

Hence, an econometrician estimating (111), will find:

$$\lambda = \frac{K}{\zeta}. \quad (113)$$

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 micro structure is done at the level of one asset, so that the λ they estimate is

$$\lambda = \frac{K}{\zeta^\perp}, \quad (114)$$

where λ^\perp is the micro-elasticity of Section G.3.

⁹¹Indeed, ε_t creates an innovation to the cumulative flow f_{t+h} equal to

$$\mathbb{E}_t [f_{t+h} - f_{t-1}] = \mathbb{E}_t [\Delta f_t + \Delta f_{t+1} + \dots + \Delta f_{t+h}] = \varepsilon_t \left(1 + (1 - \phi_g) + \dots + (1 - \phi_g)^h \right) = \frac{\varepsilon_t}{\phi_g} \left(1 - (1 - \phi_g)^{h+1} \right).$$

⁹²More 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 $\varepsilon_t := \Delta f_t - \mathbb{E}_{t-1} [\Delta f_t]$, $K = \frac{\partial}{\partial \varepsilon_t} \lim_{h \rightarrow \infty} \mathbb{E}_t [f_{t+h} - f_{t-1} | \varepsilon_t]$. For instance, if we have (35)

$$\Delta f_t = \sum_{l=1}^k a_l \Delta f_{t-l} + \varepsilon_t,$$

then the total innovation is $K = \frac{1}{1 - \sum_{l=1}^k a_l}$. We use this in Section 4.2.

⁹³Away 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 (19) with $q_t = 0$) 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 G.12.2), with $\kappa = \kappa / \Delta t \rightarrow \infty$, and ζ constant (as $\delta = \delta \Delta t$, $\delta \kappa$ is constant as $\Delta t \rightarrow 0$), we get $B \rightarrow \frac{1}{\zeta \phi_g} = \frac{K}{\zeta}$, i.e. $\Delta p_t = \frac{K}{\zeta} \varepsilon_t$.

G.5.2 Empirical values from the micro structure literature

Frazzini et al. (2018) find that buying 2.5% of the daily volume of a stock creates a permanent price impact $\Delta p = 15\text{bp}$ (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 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 is⁹⁴

$$\lambda = \frac{\Delta p}{\Delta q} = \frac{15\text{bp}}{1\text{bp}} \simeq 15.$$

Hence, the prima facie “micro structure” price impact is $\lambda \simeq 15$.⁹⁵ 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 (“meta-order”) 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 seem surprisingly large, but it is consistent with micro structure data. Bouchaud et al. (2018) report a positive autocorrelation of the decay in the signed of trades $\varepsilon_t = \text{sign}(\Delta f_t)$, qualitatively consistent with the above model. Importantly, it is also roughly quantitatively consistent too. The empirical correlation between the signs of trades, $c(h) = \text{corr}(\varepsilon_t, \varepsilon_{t+h})$, is approximately $c(h) \simeq \frac{0.25}{h^{1/2}}$ for $h \in [1, 10^3]$, which leads to $K = 1 + \sum_{h=1}^{10^3} c(h) = 16$.⁹⁶ 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 ζ^\perp of the market, by (114)). This gives, we think, a potentially satisfactory unification of the very high impact measured in the micro structure literature, and the more moderate impact measured in inelastic markets.

One lesson is that the market micro structure 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 micro structure literature is about the micro elasticity, not the macro elasticity).

⁹⁴In 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.

⁹⁵Frazzini 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 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.

⁹⁶See 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.

G.6 The model 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

$$\pi_t = \mathbb{E}_t \left[\frac{dP_t}{P_t} \right] / dt + \frac{D_t}{P_t} - r_f \quad (115)$$

has the Taylor expansion:

$$\hat{\pi}_t = \mathbb{E}_t \frac{dp_t}{dt} + \delta (d_t - p_t). \quad (116)$$

We have:

$$q_t = -\zeta p_t + \kappa \delta d_t + \kappa \mathbb{E}_t \left[\frac{dp_t}{dt} \right] + \nu_t + f_t,$$

which in equilibrium (with $q_t = 0$) leads to the stock price equation:

$$\mathbb{E}_t \left[\frac{dp_t}{dt} \right] - \rho p_t + \delta d_t + \frac{\nu_t + f_t}{\kappa} = 0. \quad (117)$$

Integrating forward, the stock price is:

$$p_t = \mathbb{E}_t \int_t^\infty e^{-\rho(\tau-t)} \left(\frac{\rho}{\zeta} (f_\tau + \nu_\tau) + \delta d_\tau \right) d\tau \quad (118)$$

$$= \frac{\delta}{\rho} d_t + \frac{1}{\zeta} (f_t + \nu_t) + \mathbb{E}_t \int_t^\infty e^{-\rho(\tau-t)} \left(\frac{\dot{f}_\tau + \dot{\nu}_\tau}{\zeta} + \frac{\delta}{\rho} \dot{d}_\tau \right) d\tau. \quad (119)$$

This allows for easier calculations than the discrete time model. For instance, suppose that flows and dividends follow autoregressive process, e.g. $df_t = -\phi_f f_t dt + \sigma_f dZ_t$ (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 (118) gives:

$$p_t = \frac{\rho}{\rho + \phi_f} \frac{f_t}{\zeta} + \frac{\delta}{\rho + \phi_d} d_t = \frac{1}{\zeta + \kappa \phi_f} f_t + \frac{\delta \kappa}{\zeta + \kappa \phi_d} d_t. \quad (120)$$

which is the continuous time equivalent of (26).

Combining (120) and (22) leads to:

$$\hat{\pi}_t = b_f^\pi f_t + b_d^\pi d_t, \quad (121)$$

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 $b_d^\pi = \frac{\delta(1-\theta)}{\zeta+\kappa\phi_d}$.

G.7 Infinite horizon model with heterogeneous funds

When there are several funds trading stocks, the setup is very similar to the case of a single mixed fund, but with indices i for each fund. Fund i 's mandate says that the fraction invested in equities, $\frac{P_t Q_{it}}{W_{it}}$, is:

$$\frac{P_t Q_{it}}{W_{it}} = \theta_i e^{\kappa_i \hat{\pi}_t + \nu_{it}}, \quad (122)$$

where as before $\hat{\pi}_t := \pi_t - \bar{\pi}$ is the deviation of the equity premium from its average, and we allow for additional demand shocks, ν_{it} . Dividends and bond coupons are passed on to consumer, so that retained dividends would be counted as active flows.

Differentiating the demand for stocks (122) we get:

$$\Delta \ln Q_{it} = \Delta \ln W_{it} - \Delta \ln P_t + \kappa_i \Delta \hat{\pi}_t.$$

Now, from accounting the evolution of wealth of fund i is:

$$\Delta W_{it} = Q_{i,t-1} \Delta P_t + \Delta F_{it},$$

that is,

$$\frac{\Delta W_{it}}{W_{i,t-1}} = \frac{Q_{i,t-1} P_{t-1}}{W_{i,t-1} P_{t-1}} \frac{\Delta P_t}{P_{t-1}} + \frac{\Delta F_{it}}{W_{i,t-1}} = \theta_{i,t-1} \frac{\Delta P_t}{P_{t-1}} + \frac{\Delta F_{it}}{W_{i,t-1}},$$

so:

$$\Delta \ln Q_{it} = - (1 - \theta_{i,t-1}) \Delta \ln P_t + \frac{\Delta F_{it}}{W_{i,t-1}} + \kappa_i \Delta \hat{\pi}_t = - (1 - \theta_{i,t-1}) \Delta p_t + \kappa_i \Delta \hat{\pi}_t + \frac{\Delta F_{it}}{W_{i,t-1}}.$$

We use the linearization of the risk premium (18), $\hat{\pi}_t = -\delta (d_t^e - p_t) + \mathbb{E}_t [\Delta p_{t+1}]$. So, linearizing throughout,

$$\Delta \ln Q_{it} = - (1 - \theta_{i,t-1} + \kappa \delta) \Delta p_t + \kappa \Delta \mathbb{E}_t [\Delta p_{t+1}] + \Delta f_{it},$$

with:

$$\Delta f_{it} = \frac{\Delta F_{it}}{W_{it}} - (1 - \theta_{i,t-1}) \Delta \ln \bar{P}_t. \quad (123)$$

For instance, in the AR(1) model (26), we have $\mathbb{E}_t [\Delta p_{t+1}] = -\phi p_t$, so $\kappa \Delta \mathbb{E}_t [\Delta p_{t+1}] = -\kappa \phi \Delta p_t$. This definition of the flow is slightly different from that in (17): definition (123) is more useful for empirical work, and (17) for the equilibrium price. They are obviously very close. They are identical to the leading order when the baseline economy has a growth rate of zero. One can also verify directly that the two expressions are consistent in the case of the mixed fund.

G.8 Complements to the general equilibrium model

G.8.1 The model with government bonds

We propose a way to include bonds. We assume that the government issues at the beginning of period t bonds in quantities B_t . They are financed via lump-sum taxes: the tax at time t is $T_t = R_{f,t-1} B_{t-1} - B_t$, as taxes pay for the maturing bond value, $R_{f,t-1} B_{t-1}$, minus new bond issuance, B_t . We assume that the government issued debt in amount:

$$B_t = \frac{1 - \Theta}{\Theta} \bar{W}_t^\mathcal{E} \quad (124)$$

for some $\Theta > 0$, and where $\bar{W}_t^\mathcal{E} = \frac{D_t}{\delta}$ is the aggregate value of equities in the baseline model. This way, in the baseline model, total financial wealth is $\bar{W}_t^\mathcal{E} + B_t = \frac{1}{\Theta} \bar{W}_t^\mathcal{E}$, and a fraction Θ of wealth is in equities, while the rest $(1 - \Theta)$ is in government bonds. In short, the government issues enough bonds to maintain a share $1 - \Theta$ of bonds as a fraction of total financial wealth.

With this additional feature of the model, everything remains the same, except that the average equity premium is

$$\bar{\pi} = \gamma \Theta \sigma_r^2. \quad (125)$$

The calibration can be modified in a parsimonious way: we target a average aggregate equity share $\Theta = 0.6$, and increase γ by a factor $\frac{1}{\Theta}$ to keep $\bar{\pi}$ constant, changing accordingly β to keep the same interest rate r_f in the consumption Euler equation (58). Nothing else needs to change.

G.8.2 Details of the household's problem in general equilibrium: the consumer's part of the household

The consumer part of the household only decides on consumption, which entails withdrawing money from the pure bond fund. The consumption decision is made under fully rational choice: as discussed in the main text, this ensures that the standard consumption Euler equation for bonds holds, $\mathbb{E}_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{ft} \right] = 1$. Here we detail how this happens. We adopt a formulation based on mental accounts that makes this idea portable to other contexts.

The household has, at the beginning of period t , a “consumption account” with wealth w_t (this might be, for example, a checking account), and a “financial wealth account” that holds a quantity Q_t^B of bonds (with a dollar value of 1) and a quantity Q_t^E of equity shares. We let $Q_t = (Q_t^B, Q_t^E)$. While the aggregate bond holding of the household is $Q_t^B + w_t$, we posit that its two components are not fully interchangeable because of the presence of mental accounts. One can think of w_t as being small (in equilibrium, it will be 0), and Q_t^B as being big. We drop the superscript h in this section for simplicity whenever there is no ambiguity, but all the quantities refer to a given household h (which will be the representative household in equilibrium). Also, we directly refer to the ultimate assets held, as the household is assumed to be smart enough to “see through” the veil of mutual fund intermediation. Recall that Z_t is the macro state vector (see Definition 1), and we call $Z_t^h = (Z_t, Q_t^B, Q_t^E, w_t)$ the state vector specific to household h .

The evolution of wealth in the consumption account is

$$w_{t+1} = R_{ft} (w_t + \mathcal{Y}_t^h - C_t), \quad (126)$$

where

$$\mathcal{Y}_t^h = Q_t^B r_{f,t-1} + Q_t^E D_t + \Omega_t - T_t =: \mathcal{Y}(Z_t^h) \quad (127)$$

is the aggregate income to the household, coming from its “financial dividend” stream (from bonds in the financial account, $Q_t^B r_{f,t-1}$, as well as equities, $Q_t^E D_t$), plus residual income Ω_t (e.g., comprising labor income), minus T_t .

The consumer's problem is to maximize lifetime utility, subject to this dynamic budget constraint. Crucially, we assume that when the consumer does that, she takes the income \mathcal{Y}_t^h as exogenous to her consumption decisions: in this sense, the decisions of the consumer and financier sides of the households are decoupled. So, as in any consumption-saving problem, the consumer's Euler equation for bonds holds, (47). But the consumer does not see that as she consumes more, and hence lowers the amount of bonds in the household's holdings, she will induce a (typically small) flow from stocks to bonds in the financier side of the household, so that Q_{t+1} is affected. In equilibrium, $w_t = 0$ and $\mathcal{Y}_t^h = Y_t = C_t$ at all dates, as the representative household consumes aggregate income.

It might be helpful to also state this same argument using dynamic programming. Namely, the consumer side of the household solves:

$$V(w_t, Z_t^h) = \max_{C_t} u(C_t) + \beta \mathbb{E} [V(R_{ft}(w_t + \mathcal{Y}(Z_t^h) - C_t), Z_{t+1}^h)], \quad (128)$$

where the law of motion of Z_{t+1}^h is taken by the agent to be independent of c_t . In the case of a strategic consumer wishing to manipulate the financier part of the household, the agent would take into account the (small) dependence on Z_{t+1}^h on c_t .

G.8.3 Details of the household's problem in general equilibrium: mutual fund shares accounting

This section provides some extra details on the household's problem of Section 5.1. We call t^- the beginning of period values, evaluated at the time t price P_t . The mixed fund gives a dividend $D_t^M = QD_t + r_{f,t-1}B_{t-1}^M$, so that its cum-dividend value is $W_{t^-}^M = QP_t + B_{t-1}^M + D_t^M$, 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 $v_t^M = \frac{QP_t + B_{t-1}^M}{N_{t-1}}$. So, the beginning of period wealth of the household is:

$$W_{t^-}^h = \frac{N_{t-1}^h}{N_{t-1}} W_{t^-}^M + B_{t-1}^h R_{f,t-1}. \quad (129)$$

Suppose there are household flows ΔF_t^h into the mixed fund, while flows from the rest of the economy are ΔF_t (in equilibrium, the two values are the same). Then, the number of shares owned by the household and in the fund are:⁹⁷ $N_t^h = N_{t-1}^h + \frac{\Delta F_t^h}{v_t^M}$ 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:

$$V(W_{t^-}, Z_t) = \max_{C_t, B_t^h} u(C_t) + \beta \mathbb{E} [V(W_{t+1}^h, Z_{t+1})].$$

This problem defines a consumption policy, and also desired holdings in the pure bond fund (hence, a flow out of the bond fund).

G.8.4 On a production economy

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:

$$V(K_t, Z_t) = \max_{I_t, L_t} \{F(K_t, L_t, Z_t) - w(Z_t)L_t - I_t - \kappa(I_t, K_t, Z_t) + \mathbb{E}_t[\mathcal{M}_{t+1}V((1-\delta)K_t + I_t, Z_{t+1})]\}, \quad (130)$$

⁹⁷We also have $v_t^M = \frac{W_{t^-}^M}{N_t} = \frac{QP_t + B_{t-1}^M}{N_{t-1}}$, $B_t^M = B_{t-1}^M + \Delta F_t$. Flows change the number of shares issued by the fund, but not (controlling for stock prices) the value of each fund share.

using the SDF \mathcal{M}_{t+1} developed in the paper (see (59)). 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.

G.9 Linking flows to beliefs

G.9.1 Flows and perceived risk premia

We now develop a model in which flows are determined by households' expectations of excess returns. This simple notion turns out to be quite rich.

We will use the following notations. We call Θ_t^h the equity share of household h , and Θ_t the aggregate equity share in the economy, and Θ its baseline value. We call $\hat{\pi}_t^h$ the perception by household h of the excess risk premium $\hat{\pi}_t = \pi_t - \bar{\pi}$, and $\hat{\pi}_t^H$ the perception of this excess risk premium by the representative household. Those perceptions need not be rational. Suppose that household h has an inflow f_t^h into the market, while the average household has an inflow f_t . We simply consider a small, hypothetical divergence between household h and the representative household that happened just before t . We start with with a bit of equity-share accounting.

Lemma 1. *The equity share of household h is:*

$$\Theta_t^h = \Theta (1 + f_t^h - f_t + (1 - \Theta)p_t). \quad (131)$$

Proof. Let us call Θ_t the aggregate equity share (this is also θ_{Wt} in (15), but here we use a slightly simpler notation), using (124) for the aggregate bond supply:

$$\Theta_t = \frac{P_t Q}{P_t Q + B_t} = \frac{\bar{P}_t Q e^{p_t}}{\bar{P}_t Q e^{p_t} + \frac{1-\Theta}{\Theta} \bar{P}_t Q} = \frac{\Theta e^{p_t}}{\Theta e^{p_t} + 1 - \Theta},$$

so that, taking a first order expansion in small p_t ,

$$\Theta_t = \Theta (1 + (1 - \Theta)p_t). \quad (132)$$

Household h has an equity share $\Theta_t^h = \frac{P_t Q e^{f_t^h - f_t}}{P_t Q + B_t}$, as it has the same aggregate wealth as the representative household (in the denominator), i.e., as a Taylor expansion:

$$\Theta_t^h = \Theta_t (1 + f_t^h - f_t). \quad (133)$$

Combining those two expressions gives (131). \square

Let us now discuss the modeling of the flow f_t^h of a household h that perceives the equity premium to be $\hat{\pi}_t^h$. One natural model is that the financier optimizes over Θ_t^h (with $V^p(W) = \frac{W^{1-\gamma}-1}{1-\gamma}$) the “narrow framing” value function:

$$U(\Theta_t^h, \hat{\pi}_t^h) := \mathbb{E} [V^p(R_{f_t} + \Theta_t(\bar{\pi} + \hat{\pi}_t^h + \varepsilon_{t+1}^r))] \simeq 1 + r_{f_t} + \Theta_t^h(\bar{\pi} + \hat{\pi}_t^h) - \frac{1}{2}\gamma\sigma_r^2(\Theta_t^h)^2, \quad (134)$$

where ε_{t+1}^r has mean 0 and variance σ_r^2 , and the second expression is in the limit of small time intervals. This implies a targeted equity share (using $\pi_t^h = \bar{\pi} + \hat{\pi}_t^h$) equal to:

$$\Theta_t^h = \frac{\pi_t^h}{\gamma\sigma_r^2} = \frac{\bar{\pi} (1 + \kappa^H \hat{\pi}_t^h)}{\gamma\sigma_r^2},$$

with $\kappa^H = \frac{1}{\bar{\pi}}$.

This “attentive portfolio choice” formulation, while it may be appealing as it is traditional, leads to three problems. First, in the cross-section, the attentive formulation predicts:

$$f_t^h - f_t = \kappa^H (\hat{\pi}_t^h - \hat{\pi}_t^H), \quad (135)$$

with a strength of $\kappa^H = \frac{1}{\bar{\pi}} \simeq 22$. This is a very large pass-through from beliefs to actions, and it contradicts both intuition (as an agent seeing a higher equity premium by 2.5% would double his equity share), and the evidence in Giglio et al. (2021a), who instead find a passthrough $\kappa^H \simeq 2$ (using yearly units). The second problem is that, with this formulation, we cannot model a plausible agent who would let his equity share drift passively.⁹⁸ The equity share Θ_t^h is always “actively managed” by the household. The third problem is that the household is so active that it exactly cancels out all the actions of intermediaries, i.e. of the mixed fund. Indeed, plugging this into (132) gives a price equal to $p_t = \frac{\kappa^H \hat{\pi}_t^H}{(1-\Theta)}$, so proportional to the risk premium $\hat{\pi}_t^H$ perceived by the representative household, and independent of ζ and any other considerations.⁹⁹

We propose a common resolution of those three problems, which relies on “behaviorally inattentive” portfolio choice. We posit that a household h is partially inattentive to some of the normatively relevant determinants of its flow, and chooses its flow according to:

$$\max_{f_t^h} U(\Theta^h(f_t^h, m_\theta f_t, m_\pi p_t), m_\pi \hat{\pi}_t^h), \quad (136)$$

where $\Theta^h(f_t^h, f_t, p_t) = \Theta(1 + f_t^h - f_t + (1 - \Theta)p_t)$ as in (131), U is in (134), $m_\theta \in [0, 1]$ is the attention to the rebalancing concerns, and m_π is the attention to one’s estimate of the risk premium. The traditional case corresponds to m_θ and m_π equal to 1. However, when $m_\theta = 0$, the household pays no attention to rebalancing needs: it lets the portfolio drift. This sort of behavioral modeling is tractable, it applies to a variety of micro and macro setups, and has good empirical support (see the survey in Gabaix (2019)). We derive the resulting flows.

Proposition 13. *In the above behavioral formulation, the aggregate flow is linked to the aggregate belief $\hat{\pi}_t^H$ by:*

$$f_t = \frac{\kappa^H \hat{\pi}_t^H - m_\theta (1 - \Theta) p_t}{1 - m_\theta}, \quad (137)$$

where:

$$\kappa^H = \frac{m_\pi}{\bar{\pi}}. \quad (138)$$

⁹⁸For instance, Giglio et al. (2021b) document a large fraction of investors who do not trade at all during the stock market crash of early 2020.

⁹⁹In a standard model with two types of agents, both agents matter. However, here the household is the principal, and the mixed fund the agent, so that the household in fact owns all equities, and thus (if it is active enough) it can undo the actions of the mixed fund.

In the cross-section, a household h perceiving the equity premium deviation to be $\hat{\pi}_t^h$ has a flow:

$$f_t^h = f_t + \kappa^H (\hat{\pi}_t^h - \hat{\pi}_t^H), \quad (139)$$

so that the cross-sectional sensitivity of flows to the equity premium is κ^H , less than then the aggregate sensitivity of flows to the equity premium is, which is $\frac{\kappa^H}{1-m_\theta} > \kappa^H$. Those two relations also imply the following link:

$$f_t^h = m_\theta (f_t - (1 - \Theta)p_t) + \kappa^H \hat{\pi}_t^h. \quad (140)$$

Proof. Given (136), at the optimal flow level we should have, with $\kappa^H = \frac{m_\pi}{\pi}$:

$$\Theta_t^h = \frac{\pi_t + m_\pi \hat{\pi}_t^h}{\gamma \sigma^2} = \frac{\bar{\pi} (1 + \kappa^H \hat{\pi}_t^h)}{\gamma \sigma^2} = \Theta (1 + \kappa^H \hat{\pi}_t^h).$$

We also have

$$\Theta_t^h = \Theta^h (f_t^h, m_\theta f_t, m_\theta p_t) = \Theta (1 + f_t^h - m_\theta f_t + (1 - \Theta) m_\theta p_t).$$

This gives (140). Finally, the flow of the representative agent satisfies $f_t^h = f_t$, i.e. is (137). \square

We check that this formulation solves the three problems we outlined. First, the cross-household sensitivity of flows to beliefs is smaller, and can be realistically calibrated. Intuitively, the household does see that high perceived excess return $\hat{\pi}_t^h$ should lead to a high flow f_t^h ($m_\pi > 0$), but doesn't agree that a 2.5% extra risk premium should lead it to increase its equity share from $\Theta = 60\%$ to 94%, as it would in the rational model (with $\kappa^H = \frac{1}{\pi} = 22$). Indeed, Giglio et al. (2021a) find $\kappa^H \simeq 2$.¹⁰⁰ To match their evidence, we need to calibrate $m_\pi = \kappa^H \bar{\pi} = 0.08$. This means that people might have “bold forecasts” but “timid choices”, very much as in Kahneman and Lovallo (1993). Second, we can model a very inactive household that would let its wealth drift independently of any perception of the risk premium: that would correspond to the case in which $m_\theta = m_\pi = 0$. Third, now the household does not exactly cancel the actions of the mixed fund: in (137), because $m_\theta < 1$, the flow's reaction is finite, and only in the completely reactive case ($m_\theta = 1$) the household would completely undo the actions of the mixed fund (when $m_\theta \rightarrow 1$, to avoid an infinite flow the numerator of (137) needs to go to 0, which pins the price to $p_t = \frac{\kappa^H \hat{\pi}_t^H}{1-\Theta}$).

G.9.2 Application

Flows and expected risk premia: A simple specification As a simple calibration of (137), let us take $m_\theta = 0$. Then, we have:

$$f_t = \kappa^H \hat{\pi}_t^H, \quad (141)$$

i.e. the flow is just the perceived risk premium, times the sensitivity κ^H . If $\kappa^H = 2$, then the volatility of the perceived risk premium is simply the volatility of the flows, divided by $\kappa^H = 2$, so about 1.4% per year in our calibration (see Section 5.4). Recall that we have $f_t = \theta b_t$ (see (51)). So, this is a model where the “behavioral disturbance” is simply a time-variation in the perceived value of equities, $b_t = \frac{\kappa^H}{\theta} \hat{\pi}_t^h$

¹⁰⁰For instance, column 6 of their Table III gives $\frac{d\Theta_t^h}{d\pi_t^h} = 1.2$. Given a typical $\Theta = 2/3$, that leads to $\kappa^H = \frac{1}{\Theta} \frac{d\Theta_t^h}{d\pi_t^h} = 1.8$.

Flows and expected growth of fundamentals Bordalo et al. (2020) propose that stock market fluctuations are linked to subjective expectations of long-term growth in dividends. Let us see how that would be linked to flows and prices in inelastic markets. Call g_t the expectation of the growth rate of dividends (expressed as a deviation from the average growth rate \bar{g} , so that the expected growth rate is $\bar{g} + g_t$), and for now suppose that

$$f_t = b_g^f g_t,$$

for some parameter $b_g^f \geq 0$. We assume that g_t mean-reverts with speed ϕ , so that $\mathbb{E}_t[g_{t+1}] = (1 - \phi)g_t$. Then, the price deviation from the baseline is:¹⁰¹

$$p_t = \frac{\kappa + b_g^f}{\zeta^M} g_t, \quad (142)$$

with $\zeta^M = \zeta + \kappa\phi$. In a frictionless economy where agents hold those beliefs, the price would be $p_t^* = \frac{1}{\delta + \phi} g_t$. So the flow passthrough that replicates that is $b_g^{f,*} = \frac{\zeta^M}{\delta + \phi} - \kappa$. Calibrating, this gives $b_g^{f,*} = \frac{0.2}{0.034 + 0.04} - 1 \simeq 2$ years. If the growth rate is perceived to be 1% higher, then the flow is 2% higher.

Now, let us see how we link the flows more deeply to explicit beliefs about returns. We find it useful to consider that households have time-varying beliefs about the long-term growth rate of dividends, $m_g g_t$, and the representative fund has beliefs g_t . When $m_g > 1$, the general public is more reactive in its beliefs than institutions.

Proposition 14. *The solution is as follows. Flows, prices, and the risk premium perceived by institutions ($\hat{\pi}_t$) and households ($\hat{\pi}_t^H$) follow:*

$$f_t = b_g^f g_t, \quad p_t = b_g^p g_t, \quad \hat{\pi}_t = b_g^\pi g_t, \quad \hat{\pi}_t^H = b_g^{\pi,H} g_t,$$

where, with $\zeta^M = \zeta + \kappa\phi$,

$$b_g^p = \frac{m_g \kappa^H + (1 - m_\theta) \kappa}{\kappa^H (\delta + \phi) + (1 - m_\theta) \zeta^M + m_\theta (1 - \Theta)},$$

and $b_g^f = \zeta^M b_g^p - \kappa$, $b_g^\pi = 1 - (\delta + \phi) b_g^p$, $b_g^{\pi,H} = m_g - (\delta + \phi) b_g^p$.

Proof. The risk premium perceived by institutions is the belief $\hat{\pi}_t^I = g_t - (\delta + \phi) p_t$, so

$$b_g^\pi = 1 - (\delta + \phi) b_g^p, \quad (143)$$

while the belief of households is $\hat{\pi}_t^H = m_g g_t - (\delta + \phi) p_t$, so that its sensitivity to g_t is $b_g^{\pi,H} = m_g - (\delta + \phi) b_g^p$.

¹⁰¹Indeed, replacing g by $\bar{g} + g_t$ in (74) we obtain (in the limit of small time intervals):

$$\hat{\pi}_t = g_t + \mathbb{E}_t[\Delta p_{t+1} + \delta(d_{t+1} - p_t)].$$

We seek a solution of the type $p_t = b_g^p g_t$, and we also observe that the deviation of the dividend from the baseline is simply $d_{t+1} = 0$. Plugging this in (65), setting $\nu_t = 0$ for simplicity, gives:

$$0 = q_t = - (1 - \theta) p_t + \kappa \hat{\pi}_t + f_t = - (1 - \theta) b_g^p g_t + \kappa (g_t - (\phi + \delta) b_g^p g_t) + b_g^f g_t,$$

which gives (142).

By market clearing in the equity market, the total demand change by the mixed fund should be 0:

$$0 = q = -(1 - \theta) p_t + f_t + \kappa \hat{\pi}_t^I = -(1 - \theta) p_t + f_t + \kappa (g_t - (\delta + \phi) p_t),$$

i.e., with $\zeta^M = 1 - \theta + (\delta + \phi) \kappa = \zeta + \kappa \phi$,

$$0 = -\zeta^M b_g^p + b_g^f + \kappa,$$

hence we recover (142),

$$b_g^f = \zeta^M b_g^p - \kappa \tag{144}$$

Finally, the flow of the representative investor is as in (137), so that:

$$0 = (1 - m_\theta) f_t + (1 - \Theta) m_\theta p_t - \kappa^H \hat{\pi}_t^H,$$

so

$$\begin{aligned} 0 &= (1 - m_\theta) b_g^f + m_\theta (1 - \Theta) b_g^p - \kappa^H b_g^{\pi, H} \\ &= (1 - m_\theta) (\zeta^M b_g^p - \kappa) + m_\theta (1 - \Theta) b_g^p - \kappa^H (m_g - (\delta + \phi) b_g^p), \end{aligned}$$

and

$$b_g^p = \frac{m_g \kappa^H + (1 - m_\theta) \kappa}{\kappa^H (\delta + \phi) + (1 - m_\theta) \zeta^M + m_\theta (1 - \Theta)}. \tag{145}$$

□

An illustrative calibration To match the low passthrough from beliefs to actions in Giglio et al. (2021a), $\kappa^H \simeq 2$ years, we use $m_\pi = \kappa^H \bar{\pi} = 0.08$. We set the attention to rebalancing to $m_\theta = 0.3$, and give equal beliefs to households and institutions. This leads to $b_g^\pi = 0.5$. This means that if agents perceive the growth rate to be 1 percentage point higher than usual, they think that the risk premium is only 0.5 percentage points higher: they believe that the market has incorporated half of those news. This is also in line with Giglio et al. (2021a) (Table IX). This implies that $b_g^p = 6.6$ and $b_g^f = 0.3$. This calibration is illustrative. In future research, it would be highly desirable to quantify those parameters directly, in particular the only partial tendency of households to rebalance ($m_\theta < 1$).

G.10 Pricing kernel consistent with flow-based pricing: Complements

This section gives complements to the flow-based SDF of Section 5.3.

G.10.1 Basics

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: $\mathcal{M}^d = \frac{1}{R_f}$.¹⁰²

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 will not impact prices) of “agile optimizers,” who start with zero financial wealth and whose objective function is:

$$\max_Q \mathbb{E} \left[-\mathcal{M}^d e^{-Q'R} \right], \quad (146)$$

where R is the vector of excess returns at time 1.¹⁰³ 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 .¹⁰⁴ Hence we have $\mathbb{E} [\mathcal{M}^d e^{-Q'R} R] = 0$. So, the following \mathcal{M} is a pricing kernel:

$$\mathcal{M} = \mathcal{M}^d e^{-Q'R + \xi}, \quad (147)$$

where the constant ξ ensures that the risk-free rate is correctly priced ($\mathbb{E}[\mathcal{M}] = \mathbb{E}[\mathcal{M}^d]$, so $\xi = \ln \frac{\mathbb{E}[\mathcal{M}^d]}{\mathbb{E}[\mathcal{M}^d e^{-Q'R}]}$).

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 (147) 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

$$p = \frac{f + \kappa \delta d}{\zeta}, \quad (148)$$

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:

$$\mathcal{M} = \exp \left(-r_f - \pi \frac{\varepsilon^D}{\sigma_d^2} + \xi \right), \quad \pi = \bar{\pi} - \delta \frac{\tilde{f}}{\zeta}, \quad (149)$$

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 (149) just reflects that. If there is a flow f , that modifies the price P according to (148), 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.

¹⁰²In 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 G.10.2.

¹⁰³The implicit risk aversion of 1 is just a normalization.

¹⁰⁴They start with zero capital at each period, and rebate their profits and losses to the representative household.

Flow-based SDF for the infinite-horizon model Section 5.3 developed the SDF for the infinite-horizon model process for flows in (26), something also delivered by our general equilibrium model of Section 5. 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.

G.10.2 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 We start with a basic example. We suppose that returns and consumption are lognormal:

$$\frac{C_1}{C_0} = e^{g_c + \sigma_c \varepsilon^c - \frac{1}{2} \sigma_c^2}, \quad (150)$$

with ε_t^c a standard Gaussian random variable. Consider the consumption pricing kernel, which is:

$$\mathcal{M}^{d,C} \equiv e^{M^{d,C}} = \beta \left(\frac{C_1}{C_0} \right)^{-\gamma} = e^{-r_f - \gamma \sigma_c \varepsilon^c - \frac{1}{2} \gamma^2 \sigma_c^2}$$

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:

$$\mathcal{M}^d = e^{M^d},$$

which might be \mathcal{M}^{d,R_f} or $\mathcal{M}^{d,C}$. We recall that for two jointly Gaussian variables X, Y :

$$\frac{\mathbb{E}[e^X Y]}{\mathbb{E}[e^X]} = \mathbb{E}[Y] + \text{cov}(X, Y). \quad (151)$$

For instance, the anomalous excess equity premium is

$$\mathbb{E}^{\mathcal{M}^d}[R] := \frac{\mathbb{E}[\mathcal{M}^d R]}{\mathbb{E}[\mathcal{M}^d]} = \mathbb{E}[R] + \text{cov}(M^d, R), \quad (152)$$

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 (146) is $\mathbb{E}[\mathcal{M}^d e^{-Q'R} R] = 0$, so that using (151), with $V_R = \text{cov}(R, R)$ the variance-covariance matrix of returns,

$$\mathbb{E}[R] + \text{cov}(M^d, R) - V_R Q = 0,$$

where $-cov(M^d, R)$ 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:

$$R^\tau = Q' R = \mathbb{E}^{\mathcal{M}^d} [R]' V_R^{-1} R, \quad (153)$$

which depends on the “anomalous” excess returns $\mathbb{E}^{\mathcal{M}^d} [R]$. Their “excess Sharpe ratio” is:

$$\mathcal{S} = \frac{\mathbb{E}^{\mathcal{M}^d} [R^\tau]}{\sigma_{R^\tau}}, \quad (154)$$

which is the Sharpe ratio they get in excess of the average returns warranted by the default pricing kernel. Given that $\mathbb{E}^{\mathcal{M}^d} [R^\tau] = \mathbb{E}^{\mathcal{M}^d} [R]' V_R^{-1} \mathbb{E}^{\mathcal{M}^d} [R] = \sigma_{R^\tau}^2$, 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:

$$\mathcal{M} = \mathcal{M}^d e^{-\mathcal{S} \frac{R^\tau - \mathbb{E}^{\mathcal{M}^d} [R^\tau]}{\sigma_{R^\tau}} - \frac{1}{2} \mathcal{S}^2}. \quad (155)$$

This SDF \mathcal{M} prices all assets correctly: $P_a = \mathbb{E} [\mathcal{M} D_a]$ for all assets.

G.11 Corporate finance in inelastic markets: Complements

We provide complements to Section 6.2.

G.11.1 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 15. *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 $\delta = r_f + \pi - g$ to $\delta' = (1 - b) \delta = r_f + \pi - g'$, where $g' = g + G$ is 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 dividend-price 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 15 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 $P_t Q_t = \frac{D_t}{\delta}$. As bD_t dollars are devoted to purchasing shares each period, the number of shares follows $\dot{Q}_t = -\frac{bD_t}{P_t} = -bQ_t\delta$, so that the new number of shares is:

$$Q'_t = Q_0 e^{-Gt}, \quad G = b\delta. \quad (156)$$

Hence, the dividend per share in the new regime is $D'_t = \frac{(1-b)D_t}{Q'_t} = \frac{(1-b)D_t}{Q_0 e^{-Gt}}$ i.e. as $D_t = \frac{D_t}{Q_0}$,

$$D'_t = (1-b)e^{Gt} D_t. \quad (157)$$

Because the value of the firm is constant, $P'_t Q'_t = P_t Q_0$, we have

$$P'_t = P_t e^{Gt}. \quad (158)$$

This implies that the new dividend-price ratio is

$$\frac{D'_t}{P'_t} = \frac{D_t}{P_t} (1-b). \quad (159)$$

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 (156) in the last equation. □

Buybacks in an inelastic model We next study the situation in an inelastic model. We need a stabilizing force, and we assume that the flow has a “reaction to the risk premium” component $\chi \hat{\pi}_t$ as in (162).

Proposition 16. (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: a fraction b of the aggregate payout is devoted to share buybacks. Then the firm value increases by $v_* = \frac{(\mu^D - \mu^G)\theta b}{\chi}$ in the long run, and the risk premium is lower by $\hat{\pi}_* = -\delta v_*$.*

Hence, the economics is similar to the simple model of Proposition 7.

Proof. The proof is similar to that of Proposition 7 in Section F.¹⁰⁵ As in the rational case, growth rate in the number of shares is $-G$ with $G = \delta b$, $d_t = -b + Gt$. As each period the aggregate dividends are lower by b , and they represented a fraction $\delta\theta$ of the fund’s value, and capital gains are increased by G , the inflow each period, from the households to the mixed fund, is:

$$\frac{d}{dt} f_t = (1 - \mu^D) (-b\delta\theta) - \mu^G G\theta + \chi \hat{\pi}_* = (-1 + \mu^D - \mu^G) \theta G + \chi \hat{\pi}_*$$

We conjecture a long term equilibrium where $\frac{dp_t}{dt} = G$, of the type: $p_t = Gt + \tilde{p}_t$, where $\tilde{p}_t = o(t)$. The demand deviation by the mixed fund is still $q_t = -\zeta p_t + \kappa \delta d_t + f_t$, and it should be equal to the supply deviation, $q_t^S = -Gt$. This gives,

$$0 = \lim_{t \rightarrow \infty} \frac{d}{dt} (q_t - q_t^S) = -\zeta G + \kappa \delta G + (-1 + \mu^D - \mu^G) \theta G + \chi \hat{\pi}_* + G = (\mu^D - \mu^G) \theta G + \chi \hat{\pi}_*.$$

¹⁰⁵Our deviation from the baseline are in logs, e.g. $p_t = \ln \frac{P_t}{P_t}$.

This implies that the long run change in the risk premium is

$$\hat{\pi}_* = -\frac{(\mu^D - \mu^G)\theta G}{\chi} = -\frac{(\mu^D - \mu^G)\theta\delta}{\chi}b.$$

The corresponding increase in market value if $v_* = -\frac{\hat{\pi}_*}{\delta}$. □

In the long run, it is very likely that the $\mu^D - \mu^G$ will be more rational, and fall to a value closer to 0—perhaps because the financier will more actively target a constant equity share (as in Section G.9), which would create another corrective inflow. Examining this effect empirically would be interesting.

G.11.2 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:

$$q^S = -b, \quad d = 0.$$

In a frictionless model, this buyback does not change the current price per share, and does not change the time-0 return r

$$\text{Frictionless model: } p = 0, \quad v = -b, \quad r = 0.$$

In an inelastic model, now $q = -\zeta p + f^h = q^S = -b$, so

$$p = \frac{b + f^h}{\zeta}, \quad v = p - b, \quad r = p.$$

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:

$$p = \frac{b}{\zeta + \theta\mu^G}. \tag{160}$$

Infinite horizon. We suppose that the debt will be repaid very far in the future (at date $T \rightarrow \infty$). Then, the economics is as in the two-period model.¹⁰⁶

¹⁰⁶In 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.

G.12 When flows respond to risk premia in the long run

G.12.1 Basic version

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. For clarity, we use the detrending procedure of Section 3.2, but with the baseline of $(\bar{P}_t, \bar{D}_t, \bar{W}_t) = (\bar{P}_0, \bar{D}_0, \bar{W}_0) \mathcal{G}_t$ with $\mathcal{G}_t = 1$. We also use log deviations, e.g., $d_t = \ln \frac{D_t}{D_0}$ is the deviation of the dividend from the initial date 0, and is typically non-stationary, for instance we might have $\mathbb{E}_0 [d_t] = gt$. This allows us to zoom in on the core issue of how households respond to trends in d_t . This also means that $\bar{F}_t = 0$, and that f_t is the scaled flow.

Lemma 2. *(Trend-cycle decomposition for flows) The price-dividend ratio is stationary if and only if the cumulative flow f_t admits the decomposition*

$$f_t = (1 - \theta) d_t + \hat{f}_t, \quad (161)$$

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 that $q = f_t - (1 - \theta) p_t + \kappa \hat{\pi}_t$. As $q = 0$, (161) holds, with $\hat{f}_t = -\kappa \hat{\pi}_t + (1 - \theta) (p_t - d_t)$. \square

How does the market equilibrate in the long run? One might ask, how does the market discover the trend $(1 - \theta) d_t$ in (161)? 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 G.12.2 provides proof and complements. Consider a behavioral rule of the type

$$\Delta f_t = \chi \hat{\pi}_t + \varepsilon_t, \quad (162)$$

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 2, and hence the correct representation (161).¹⁰⁷

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.

¹⁰⁷Rule (162) can have microfoundations of the “behavioral inattention” type:

$$f_t = m f_t^r + (1 - m) f_{t-1} + \tilde{f}_t, \quad (163)$$

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 $\mathbb{E}_t [V^p (R_{t+1})]$) is $f_t^r = f_t + \frac{\hat{\pi}_t}{\pi}$, behavior (163) generates (162) with $\chi = \frac{m}{1-m} \frac{1}{\pi} > 0$ and $\varepsilon_t = \frac{\tilde{f}_t}{1-m}$. Formulation (163) extends more easily than (162) to other contexts (Gabaix (2014)).

Proposition 17. (Equilibrium when flows respond to the equity premium in a noisy fashion) *In the limit of small time intervals, the specification of flows (162) with i.i.d. shocks ε_t generates a deviation of the price from trend equal to:*

$$\mathbb{E}_0 p_t = \frac{1}{\zeta + \kappa\phi} \mathbb{E}_0 f_t, \quad \mathbb{E}_0 f_t = (1 - \phi)^t f_0. \quad (164)$$

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 (24). This captures that flows are endogenously “digested” by the market at a rate ϕ , which is higher when χ 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 $\chi = 0.1$, we replicate a slow mean-reversion of the P/D ratio of $\phi \simeq 4\%$ per year.¹⁰⁸

One could imagine agents with other behavioral rules, or agents optimizing on the parameters χ , m , this way providing additional cross-asset predictions.¹⁰⁹ We leave that to future research.

G.12.2 When flows react to the risk premia: advanced material

Here we derive Proposition 17, and more generally explore the consequences of flows of the type:

$$\Delta f_t = \chi \hat{\pi}_t + \varepsilon_t. \quad (166)$$

We first proceed in continuous time, which is cleanest.

Continuous time For simplicity, we assume away dividend surprises. They would be easy to add back. The flows (166) are

$$df_t = \chi \hat{\pi}_t dt + \sigma dz_t. \quad (167)$$

We use the operator D ,

$$Dx_t := \frac{\mathbb{E}_t [dx_t]}{dt}. \quad (168)$$

So, $\hat{\pi}_t = -\delta p_t + \frac{\mathbb{E}_t [dp_t]}{dt}$ (see Section G.6) becomes:

$$\hat{\pi}_t = (D - \delta) p_t \quad (169)$$

and (167) gives

$$Df_t = \chi \hat{\pi}_t. \quad (170)$$

¹⁰⁸The parameter χ is unitless: in continuous time, $df_t = \chi \hat{\pi}_t dt + \sigma dz_t$.

¹⁰⁹Alternatively, consider a rule like:

$$\Delta f_t = \chi \hat{\pi}_t + \beta (d_t - p_t) + \Delta \tilde{f}_t, \quad (165)$$

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 (165) also leads to the correct form shown in Lemma 2. 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 2 the left-hand side should not be stationary.

The basic dynamic pricing equation, (66), becomes:

$$0 = -\zeta p_t + \kappa D p_t + f_t. \quad (171)$$

Differentiating once and taking time- t expectations gives:

$$0 = -\zeta D p_t + \kappa D^2 p_t + D f_t \quad (172)$$

$$= [-\zeta D + \kappa D^2 + \chi(D - \delta)] p_t \quad (173)$$

$$= H(D) p_t$$

where

$$H(x) = \kappa x^2 - (\zeta - \chi)x - \chi\delta. \quad (174)$$

The fundamental solutions of equation $H(D)p_t = 0$ are of the form $p_t = B e^{xt}$, 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:

$$\rho = \frac{\zeta - \chi + \sqrt{\Delta}}{2\kappa}, \quad \phi = \frac{-\zeta + \chi + \sqrt{\Delta}}{2\kappa}, \quad \Delta = (\zeta - \chi)^2 + 4\chi\kappa\delta. \quad (175)$$

When $\chi = 0$, $\rho = \frac{\zeta}{\kappa}$ (as in Proposition 5) and $\phi = 0$. We record that ϕ solves:

$$(\zeta + \kappa\phi)\phi = \chi(\phi + \delta) \quad (176)$$

and as the product of the two roots, $-\phi\rho$ is equal to $\frac{-\chi\delta}{\kappa}$ in (174),

$$\phi = \frac{\chi\delta}{\kappa\rho}. \quad (177)$$

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’’):

$$p_t = A y_t, \quad f_t = a y_t.$$

Plugging this Ansatz in (167) and examining the σz_t term gives:

$$a = 1.$$

Next, we have $D y_t = -\phi y_t$, so plugging this in (171) and matching coefficients gives: $0 = [-(\zeta + \kappa\phi)A + a] y_t$, i.e.

$$A = \frac{1}{\zeta + \kappa\phi}.$$

This derived Proposition 17 in continuous time. \square

One can derive other things. For instance, here is an expression for the price, expressed in discrete time for convenience.

Proposition 18. (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:*

$$p_t = \frac{f_{t-1}}{\bar{\zeta}} + \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho}{(1+\rho)^{\tau-t+1}} \left(\frac{\tilde{f}_\tau - \tilde{f}_{t-1}}{\bar{\zeta}} + M^D d_\tau^e \right) \quad (178)$$

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 (175), and the price impact of an inflow $\Delta \tilde{f}_t$ decays at a rate ϕ .

Discrete time We use the operator

$$\nabla x_t = \mathbb{E}_t [x_{t+1} - x_t]$$

so that $\nabla x_t \simeq D x_t \times \Delta t$, where D is the continuous time operator (168). 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 (166)) 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 (66) becomes:

$$0 = -(\zeta - \kappa \nabla) p_t + f_t.$$

Pre-multiplying 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

$$\tilde{H} (x) = (\kappa + \chi \omega) x^2 - (\zeta - \chi (1 - \delta \omega)) x - \chi \delta. \quad (179)$$

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:

$$p_t = A y_t, \quad f_t = a y_t. \quad (180)$$

This implies

$$\hat{\pi}_t = (\nabla - \delta) p_t = -(\phi + \delta) A y_t.$$

Plugging this in (166) gives:

$$a = -\chi \omega (\phi + \delta) A + 1.$$

Plugging the Ansatz (180) in (66) gives:

$$0 = -(\zeta + \kappa \phi) A + a.$$

Hence, we obtain: $A = \frac{a}{\zeta + \kappa \phi}$, with

$$a = \frac{1}{1 + \chi \omega \frac{\phi + \delta}{\zeta + \kappa \phi}}. \quad (181)$$

Again, formally, we obtain the continuous time limit when $\omega \rightarrow 0$.

From discrete to continuous time We draw from Section H.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 $\omega = \Delta t$, and we can write $\mathbb{E}_t[p_{t+1}] = (1 + \omega \nabla) p_t$. Hence, calling $x = \mathbf{x} \Delta t$,

$$\begin{aligned}\tilde{H}(x) &= (\kappa + \chi\omega) x^2 - (\zeta - \chi(1 - \delta\omega)) x - \chi\delta\omega \\ \tilde{H}(x) / \Delta t &= (\boldsymbol{\kappa} + \chi\Delta t) \mathbf{x}^2 - (\zeta - \chi(1 - \delta\Delta t)) \mathbf{x} - \chi\boldsymbol{\delta}\end{aligned}$$

so that indeed, $\lim_{\Delta t \rightarrow 0} \frac{\tilde{H}(x)}{\Delta t} = H(\mathbf{x})$.

Impact of a trend on dividends We prove the following.

Proposition 19. *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, $\hat{\pi}_* = -\frac{c}{\chi}$. We have $p_t = d_t + p_*$, $f_t = (1 - \theta) d_t + f_*$, with $p_* = \frac{c}{\chi\delta}$, i.e.

$$p_* = \frac{c}{\chi\delta} = \frac{c}{\kappa\phi\rho} \quad (182)$$

and $f = \zeta p_*$. For finite t , we have

$$p_t = d_t + \left(1 - \frac{\zeta}{\zeta + \kappa\phi} (1 - \phi)^t\right) p_* \quad (183)$$

so that on impact

$$p_0 = \frac{c}{(\zeta + \kappa\phi)\rho} \quad (184)$$

where $\rho, -\phi$ are the of the characteristic polynomial $H(x)$ in (174). The flows are

$$f_t = (1 - \theta) d_t + f_* (1 - (1 - \phi)^t). \quad (185)$$

We write the rule as a deviation c from the rational flow, which is $\Delta f_t = (1 - \theta) g$ by Lemma 2. 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.¹¹⁰

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 2 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 $p_* = -\frac{\hat{\pi}_*}{\delta}$.

¹¹⁰Note that it could be obtained in the case $\chi \rightarrow 0$ from Proposition 5, which gives $p_0 = \frac{c}{\zeta\rho}$ (by plugging in $f_\tau = c\tau$).

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 (173) gives

$$H(D)p_t + c = 0. \quad (186)$$

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

$$0 = -\zeta p_0 + \kappa \frac{\mathbb{E}dp_t}{dt} \Big|_{t=0} = -\zeta p_* - (\zeta + \kappa\phi) A.$$

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 (177), which gives (184).

Finally, as $0 = -\zeta p_t + \kappa Dp_t + f_t$, we have

$$f_t = \zeta p_* + (\zeta + \kappa\phi) Ae^{-\phi t} = \zeta p_* (1 - e^{-\phi t}).$$

Using our calibration $g = 2\%$, $\theta = 87.5\%$ and rule (162) with $\chi = 0.1$ we find: $\hat{\pi}_* = \frac{1-\theta}{\chi} g = 2.5\%$.

We next prove a result that synthesizes and expands on our previous results. \square

Proposition 20. *Suppose an economy with i.i.d. dividend growth $\Delta d_t = g + \varepsilon_t^d$, and consumer flows following the semi-behavioral rule:*

$$\Delta f_t = \chi \hat{\pi}_t + (1 - \theta + \gamma) \Delta d_t + c + \varepsilon_t^f \quad (187)$$

where disturbances $\varepsilon_t^d, \varepsilon_t^f$ have mean 0 and no time correlations. The rational case obtains when $\gamma, \chi, c, \text{var}(\varepsilon_t^f)$ 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 17,

$$f_t = (1 - \theta) d_t + \hat{f}_t + f_*, \quad p_t = d_t + \frac{\hat{f}_t}{\zeta + \kappa\phi} + p_* \quad \hat{f}_t = (1 - \phi) \hat{f}_{t-1} + \varepsilon_t^f + \gamma \varepsilon_t^d. \quad (188)$$

Here in (187), 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 19). The new part is really the impact of disturbances $\varepsilon_t^d, \varepsilon_t^f$: 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 5.

Proof. The terms corresponding to the non-zero trend g and c are exactly as in Proposition 19, 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 17. 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” $\gamma \varepsilon_t^d$. \square

H Calibration of the general equilibrium model of Section 5: Details

Here we provide a detailed justification of the calibration in Section 5.4.

H.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 $\phi, \delta, \rho, \hat{\pi}$ have units of $[\text{Time}]^{-1}$, their discrete time counterparts are:

$$\phi = \phi \Delta t, \quad \delta = \delta \Delta t, \quad \rho = \rho \Delta t, \quad \hat{\pi} = \hat{\pi} \Delta t, \quad (189)$$

but as κ has unit of $[\text{Time}]$ (indeed, $\kappa \hat{\pi}$ must be unitless in an expression like $\frac{PQ}{W} = \theta e^{\kappa \hat{\pi}}$, see (1)) its discrete time counterpart is:

$$\kappa = \kappa (\Delta t)^{-1}. \quad (190)$$

As it is unitless, ζ is the same in discrete and continuous time,

$$\zeta = \zeta.$$

Finally σ_f has the units of $[\text{Time}]^{-1/2}$ so

$$\sigma_f = \sigma_f (\Delta t)^{1/2}$$

and similarly for σ_d .

H.2 Calibration steps

The calibration steps are the following.¹¹¹ We express things in annualized values, but we use the correspondence in Section H.1 to go between discrete vs continuous time notions.

1. For Tables 5 and 6, first we set $\gamma = 2, g = 2\%, \sigma_y = 0.8\%, \sigma_D = 5\%, r_f = 1\%, \sigma_f = 2.8\%$. Next we impute β from r_f given (γ, g, σ_y^2) . Then we calculate $(\sigma_r^2, \bar{\pi}, \delta)$ and set $\phi_f = 4\%, \zeta^M = 0.2, \kappa = 1, \theta = 0.875$, which jointly imply $\zeta = \zeta^M - \kappa \phi_f = 0.16$. Last we calculate $(\rho, b_f^p, b_f^x, \sigma_b)$.
2. In Table 6, we use $\hat{r}_t := r_t - \mathbb{E}_{t-1}[r_t] = \varepsilon_t^D + b_f^p \varepsilon_t^f$: so, the share of the variance of excess stock returns that is due to flows is $\frac{\text{cov}(\hat{r}, b_f^p \varepsilon_t^f)}{\text{var}(\hat{r})}$. Likewise, the share of the variance of stock returns due to fundamentals is $\frac{\text{cov}(\hat{r}, \varepsilon_t^D)}{\text{var}(\hat{r})}$. The two shares add up to 1.
3. For Table 7a, the model implied mean of P/D is

$$\mathbb{E} \left[\frac{P}{D} \right] = \mathbb{E} \left[\frac{e^{p_t}}{\delta} \right] = \frac{1}{\delta} \mathbb{E} \left[\exp \left(b_f^p \tilde{f}_t \right) \right] = \frac{1}{\delta} \exp \left(\frac{1}{2} (b_f^p)^2 \frac{\sigma_f^2}{1 - (1 - \phi_f)^2} \right), \quad (191)$$

the log mean of P/D is

$$\exp(\mathbb{E} \log P/D) = \frac{1}{\delta}, \quad (192)$$

¹¹¹We thank Lingxuan Wu for performing those computations.

and the variance of $\log D/P$ is

$$\text{Var}(\log D/P) = \text{Var}(p_t) = \frac{(b_f^p)^2 \sigma_f^2}{1 - (1 - \phi_f)^2}. \quad (193)$$

Up to now, everything is in units of continuous time.

4. For Table 7b, 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 $\varepsilon_t^D \sim \mathcal{N}(0, \sigma_D^2 \cdot \Delta t)$, $\varepsilon_t^f \sim \mathcal{N}(0, \sigma_f^2 \cdot \Delta t)$ with $\Delta t = \frac{1}{12}$ and calculate the path of d_t, \tilde{f}_t from

$$d_t - d_{t-1} = g \cdot \Delta t + \varepsilon_t^D - \frac{1}{2} \sigma_D^2 \cdot \Delta t, \quad (194)$$

$$\tilde{f}_t - \tilde{f}_{t-1} = -\phi_f \Delta t \cdot \tilde{f}_{t-1} + \varepsilon_t^f, \quad (195)$$

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:

$$\log D_t = d_t, \quad (196)$$

$$p_t = b_f^p \tilde{f}_t, \quad (197)$$

$$\log P_t = \log D_t + p_t - \log \delta, \quad (198)$$

$$\log D_t/P_t = \log D_t - \log P_t, \quad (199)$$

$$\pi_t = \bar{\pi} + b_f^\pi \tilde{f}_t. \quad (200)$$

- (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.

Table I.15: Parameters and elasticities in numerical models

Variable	Lucas I	Lucas II	LRR I	IRR II
Risk aversion (γ)	4	4	10	10
IES ($\frac{1}{\varphi}$)	1/4	1/4	1.5	1.5
Time preference (β)	0.9	0.9	0.94	0.94
Mean output growth (g)	0.04	0.04	0.018	0.018
Std. dev. of output growth (σ_y)	0.1	0.1	0.046	0.046
Mean financialization ratio ($\psi = \frac{D}{Y}$)	0.2	0.2	0.082	0.082
Persistence of financialization ratio (ρ_x)	1.0	0.95	-	-
Std. dev. of financialization shocks (σ_x)	0.0	0.05	-	-
Growth rate persistence (ρ)	-	-	0.044	0.044
Growth risk loading (φ_e)	-	-	0.044	0.044
Stochastic volatility (σ_w)	-	-	0.0	0.8×10^{-5}
Persistence of volatility (ν_1)	-	-	0.854	0.854
Macro elasticity (ζ)	20.0	17.5	28.85	28.96

Notes. All parameters in the table are annualized. The parameters are calibrated so that in each model the risk premium is around 4% and the equity market’s capitalization is 1.25 times GDP. These parameters are defined throughout our discussion. In *Lucas I*, we reproduce the analytical results of Section F.4 using the numerical methods to confirm their consistency. In *Lucas II*, we compute the macro elasticity in an economy where labor income is partially correlated with financial market returns. *LRR I* follows the notation and parameterization of Case I in Bansal and Yaron (2004) with long run risks and constant volatility. We adjust the discount rate and output volatility in order to hit the calibration targets. *LRR II* follows the parameterization in their Case II with stochastic volatility.

I Computing elasticities numerically in equilibrium models

I.1 Summary of the main results

This section computes numerically the macro elasticity in classic asset pricing models.¹¹² This complements the closed forms that we obtained in Section F.4. We consider two important classes of models with different parameterizations: the Lucas (1978) model with labor income, and long run risk models as in Bansal and Yaron (2004).

Table I.15 summarizes the parameterization and the moments generated by the numerical exercises. These parameters are defined throughout our discussion. We find that the macro elasticity is around 17.5 in a Lucas model with partially correlated labor income, and close to 30 in long run risk models. Section I.2 describes the general procedure that we follow to compute the macro elasticities numerically, and Sections I.3 to I.4 outline the models and the numerical algorithms in detail.

¹¹²We thank Zhiyu Fu for performing those computations.

I.2 General procedure

For both models considered, we suppose $\Delta \ln Y_t = g - \frac{\sigma_Y^2}{2} + \varepsilon_{Yt}$ and define $\Omega_t = Y_t - D_t$ to be residual “labor income” (broadly understood). Suppose that stocks are tradable (with dividends D_t), but not labor income. The consumer receives a stream of labor income Ω_t every period. At time 0, he is endowed with all the shares in the “stock market.”

Call $\psi_t = \frac{D_t}{Y_t} = 1 - \frac{\Omega_t}{Y_t}$ the relative size of dividends versus aggregate consumption, to which we refer as the degree of “financialization” of the economy. In Bansal and Yaron (2004), we assume that ψ_t is a constant at all dates. In the Lucas model, we allow ψ_t to be time-varying so that labor income is only partially correlated with the stock market.

I.2.1 General procedure

Define Z_t to be a vector of aggregate state variables and \bar{Z} the unconditional mean of Z_t . In a Lucas model with partially-correlated labor income, $Z_t = \psi_t$, and in a long run risk model $Z_t = (x_t, \sigma_t^2)'$. We compute the equilibrium price-dividend ratio and risk-free rate, $\frac{P_t^*}{D_t} = pd^*(Z_t)$ and $r_t^{f*} = r^{f*}(Z_t)$. We define the perturbed price-dividend ratio as:

$$pd(Z_t, p) = pd^*(Z_t)(1 + p)$$

for a small constant p close to 0.

Given $(pd(Z_t, p), r^{f*}(Z_t))$, we compute the value function, $V(Z_t, w_t; p)$, the optimal portfolio rule $\theta(Z_t, w_t; p)$, and the optimal consumption plan $c(Z_t, w_t; p)$ for households with different w_t , where $w_t = W_t^\varepsilon / Y_t$ is their financial wealth relative to GDP, which equals aggregate consumption. The representative household holds $\bar{w}(Z_t) = \psi(Z_t)(pd^*(Z_t) + 1)$ in financial wealth relative to GDP. We calibrate the parameters so that on average the consumption/market capitalization is equal to $\frac{Y}{W_t^\varepsilon} = \frac{1}{\bar{w}(Z)} = 0.8$. Define $\theta^*(Z_t, p) \equiv \theta(Z_t, \bar{w}(Z_t), p)$ and $c^*(Z_t, p) = c(Z_t, \bar{w}(Z_t), p)$ as the policy functions for the representative household given the perturbation p . We have $\theta^*(Z, 0) = 1$ and $c^*(Z, 0) = 1$ for any Z .

The macro elasticity is then give as:

$$\zeta^r = -\frac{dQ^\varepsilon / Q^\varepsilon}{dP / P} = -\frac{d\theta^*(\bar{Z}, p)}{dp} \Big|_{p=0}$$

The numerical solution procedure is as follows:

1. We solve for the equilibrium objects $pd^*(Z_t)$ and $r^{f*}(Z_t)$ on the grid of aggregate state variables Z_t using standard methods, as discussed below.
2. Taking the unperturbed $pd^*(Z_t)$, $r^{f*}(Z_t)$ as given, we solve for the optimal policy functions via the following dynamic programming problem:

$$V_t(Z_t, w_t) = \max_{(\theta, c)} u(c, Z_t) + \beta \mathbb{E}_t \left[\left(\frac{Y_{t+1}}{Y_t} \right)^{1-\gamma} V(Z_{t+1}, w_{t+1}) \right],$$

where¹¹³:

$$w_{t+1} = \frac{Y_t}{Y_{t+1}} (w_t + 1 - \psi_t - c_t) R_{t+1},$$

with R_{t+1} the gross return on financial assets:

$$R_{t+1} = \theta_t \frac{D_{t+1}}{D_t} \frac{pd(Z_{t+1}) + 1}{pd(Z_t)} + (1 - \theta_t) \left(1 + r_t^{f*}(Z_t) \right).$$

For the long run risk model, the wealth evolution is the same, but the Bellman equation uses Epstein-Zin preferences, so it is nonlinear. We exploit the homotheticity of the long run risk model to speed up the computation, as explained below.

3. We check the consistency of the policy functions with the aggregate equilibrium, $\theta^*(Z, 0) = 1$ and $c^*(Z, 0) = 1$.
4. We solve another dynamic programming with the perturbed price-dividend ratio, taking $p = 0.01$. We then compute:

$$\zeta^r = -\frac{\theta^*(\bar{Z}, 0.01) - \theta^*(\bar{Z}, 0)}{0.01}.$$

I.3 Lucas models with partially correlated labor income

We allow for shocks to the financialization ratio ψ in the Lucas model. We consider the parameterization $\psi_t = \frac{x_t}{\mu_x + x_t}$, where $\ln x_t = \rho_x \ln x_{t-1} + \sigma_x \epsilon_{x,t}$, $\mu_x > 0$. In the simple Lucas model studied in the main body of the paper, we have $\psi_0 = \psi < 1$, but $\sigma_x = 0$.

I.3.1 Stationary equilibrium

We first solve for the steady state price-dividend ratio. The asset pricing equation is, with M_{t+1} being the SDF:

$$\begin{aligned} \mathbb{E}_t \left[\frac{M_{t+1}}{M_t} R_{t+1} \right] &= 1, \\ \mathbb{E}_t \left[\beta \left(\frac{Y_{t+1}}{Y_t} \right)^{-\gamma} \frac{D_{t+1}}{D_t} \frac{pd(\psi_{t+1}) + 1}{pd(\psi_t)} \right] &= 1, \\ \mathbb{E}_t \left[\beta \left(\frac{Y_{t+1}}{Y_t} \right)^{1-\gamma} \frac{\psi_{t+1}}{\psi_t} \frac{pd(\psi_{t+1}) + 1}{pd(\psi_t)} \right] &= 1. \end{aligned} \tag{201}$$

For ease of computation, we decompose the shock to x into $\varepsilon_{Y,t}$ and another component independent from it:

$$\ln x_t = \rho_X \ln x_{t-1} + \sigma_x \epsilon_{x,t} + \sigma_{x,Y} \varepsilon_{Y,t} \quad \epsilon_{x,t} \perp\!\!\!\perp \varepsilon_{Y,t}. \tag{202}$$

The stationary equilibrium is defined as the function $pd(\psi)$ that solves equation (201) subject to the law of motion of the underlying state variable x in (202).

¹¹³Indeed, the law of motion for financial wealth is $W_{t+1}^{\mathcal{E}} = (W_t^{\mathcal{E}} + \Omega_t - C_t) R_{t+1}$, which gives

$$w_{t+1} = \frac{W_{t+1}^{\mathcal{E}}}{Y_{t+1}} = \frac{Y_t}{Y_{t+1}} \left(\frac{W_t^{\mathcal{E}} + \Omega_t - C_t}{Y_t} \right) R_{t+1} = \frac{Y_t}{Y_{t+1}} (w_t + 1 - \psi_t - c_t) R_{t+1}.$$

Discretization The function $pd(\psi)$ is computed on a discrete grid over ψ . The upper and lower boundaries of the grid are chosen to cover at least 99.9% of the stationary distribution of ψ . Off-grid values are computed using linear interpolation or extrapolation. Expectations are computed using Gauss–Hermite quadrature with 10 nodes. Increasing the number of nodes does not change the results in a noticeable way. After discretization, the solution to (201) is essentially a vector corresponding to the P/D ratio on the grid for ψ . This equation is then solved non-linearly using Newton’s method.

The risk-free rate is the same as in the simple Lucas model, as the state variable ψ does not enter the SDF. The expected risky return can be computed as:

$$\mathbb{E}R_{t+1} = \mathbb{E}_t \left[\frac{Y_{t+1}}{Y_t} \frac{\psi_{t+1}}{\psi_t} \frac{pd(\psi_{t+1}) + 1}{pd(\psi_t)} \right].$$

I.3.2 Dynamic programming: Outline of the algorithm

The dynamic programming problem is solved using policy function iteration and the endogenous grid point method. For notational ease, for each variable x we denote its value in the next period as x' .

The dynamic programming problem can be solved in two steps. First, given post-consumption wealth $\tilde{w} = w + 1 - \psi - c$, the households solve a portfolio choice problem that maximizes:

$$\begin{aligned} & \max_{\theta} \mathbb{E} \left[\left(\frac{Y'}{Y} \right)^{1-\gamma} V \left(\psi', \frac{Y}{Y'} \tilde{w} R^*(\theta) \right) \right] \\ & \text{s.t.} \\ & R^* = \theta R + (1 - \theta) (1 + r_f). \end{aligned}$$

The maximization problem yields the first-order condition:

$$\mathbb{E} \left[\left(\frac{Y'}{Y} \right)^{-\gamma} V_w \left(\psi', \frac{Y}{Y'} \tilde{w} R^*(\theta^*) \right) (R - r_f) \right] = 0.$$

With knowledge of V_w (either from an initial guess or from the previous iteration), we can solve for θ^* non-linearly given each \tilde{w} . This way, we obtain the functions $\theta^*(\tilde{w})$ and $R^*(\tilde{w})$, which map post-consumption wealth \tilde{w} into the optimal portfolio choice $\tilde{\theta}$ and into the asset returns R^* , independently of the consumption policy.

Second, the household choose optimal consumption by maximizing:

$$\max_c u(c) + \beta \mathbb{E} \left[\left(\frac{Y'}{Y} \right)^{1-\gamma} V \left(\psi', \frac{Y}{Y'} \tilde{w} R^*(\tilde{w}) \right) \right] \quad (203)$$

s.t.

$$\tilde{w} = w + 1 - \psi - c, \quad (204)$$

which gives the first-order condition:

$$c^{-\gamma} = \beta \mathbb{E} \left[\left(\frac{Y'}{Y} \right)^{-\gamma} V_w \left(\psi', \frac{Y}{Y'} \tilde{w} R^*(\tilde{w}) \right) R^*(\tilde{w}) \right].$$

Following the endogenous grid point method, we can solve for c explicitly as a function of the post-consumption wealth \tilde{w} on the grid, that is,

$$c^*(\tilde{w}) = \left\{ \beta \mathbb{E} \left[\left(\frac{Y'}{Y} \right)^{-\gamma} V_w \left(\psi', \frac{Y}{Y'} \tilde{w} R^*(\tilde{w}) \right) R^*(\tilde{w}) \right] \right\}^{-\frac{1}{\gamma}}.$$

then we can back out pre-consumption wealth w (off grid) as a function of \tilde{w} from the budget constraint (204). Finally, using interpolation, we can get the desired policy function $c^*(w)$ on the grid.

By the envelope theorem, we have:

$$V_w = u'(c^*(w)) = (c^*)^{-\gamma}.$$

Using the solved policy function $c^*(w)$ we obtain V_w from the envelope theorem and proceed to the next iteration until we achieve the desired accuracy.

Verification Now, we verify that the policy functions are consistent with the aggregate equilibrium. For representative households, that is, households with financial wealth $w^\mathcal{E} = (pd(\psi) + 1)\psi$, their consumption is exactly equal to GDP, $c = 1$, and their risky share is also equal to one, $\theta = 1$.

Initial guess We make the initial guess by assuming the aggregate state ψ is constant. In this case, the solutions for each value of ψ are independent of one another. We guess and verify that the value function is of the form $V(w) = \frac{A^{-\gamma} w^{1-\gamma}}{1-\gamma}$, where w includes capitalized labor income.

I.4 Long run risk model

I.4.1 Stationary equilibrium

As shown in Bansal and Yaron (2004), the SDF for a recursive utility function is:

$$M_{t+1} = \beta^\vartheta \left(\frac{Y_{t+1}}{Y_t} \right)^{-\frac{\vartheta}{\varphi}} R_w^{\theta-1},$$

where:

$$R_w = \frac{pd(x', \sigma') + 1}{pd(x, \sigma)} \frac{Y_{t+1}}{Y_t},$$

$$\vartheta = \frac{1-\gamma}{1-\frac{1}{\varphi}}.$$

Following Bansal and Yaron (2004), the dynamics of the economy are:

$$\begin{aligned} x_{t+1} &= \rho x_t + \varphi_e \sigma_t e_{t+1}, \\ g_{t+1} &= \mu + x_t + \sigma_t \eta_{t+1}, \\ \sigma_{t+1}^2 &= \sigma^2 + v_1 (\sigma_t^2 - \sigma^2) + \sigma_w w_{t+1}, \\ e_{t+1}, u_{t+1}, \eta_{t+1}, w_{t+1} &\sim N.i.i.d.(0, 1). \end{aligned}$$

The price-dividend ratio can be solved using:

$$\mathbb{E} \left[\beta^\vartheta \left(\frac{Y_{t+1}}{Y_t} \right)^{-\frac{\vartheta}{\varphi}} R_{t+1}^\vartheta \right] = 1,$$

$$\mathbb{E} \left[\beta^\vartheta \exp((1-\gamma)g) \left(\frac{pd(x_{t+1}) + 1}{pd(x_t)} \right)^\vartheta \right] = 1.$$

Moving the current pd outside of the expectation operator:

$$pd(x, \sigma^2) = \left[\mathbb{E} \beta^\vartheta \exp((1-\gamma)g) (pd(x', \sigma^2) + 1)^\vartheta \right]^{\frac{1}{\vartheta}},$$

which gives a fixed point problem. We solve for $pd(x, \sigma^2)$ by iterating upon this equation.

I.4.2 A shortcut for the dynamic programming problem

Labor income can also be modeled as the dividend from human capital, W^L . In our setup of long run risk models, labor income is co-integrated with aggregate output, so human capital W^L and W^E have the same return process. Therefore, we could first solve the bellman equation with total wealth $W = W^L + W^E$ as the state variable, and then take care of the difference between W^L and W^E when calculating the elasticity.

The Bellman equation with Epstein-Zin preferences is:

$$V(x, w) = \max_{c, \vartheta} \left((1-\beta)c^{1-\frac{1}{\varphi}} + \beta \mathbb{E} [V^{1-\gamma}(x', w')]^{\frac{1-\frac{1}{\varphi}}{1-\gamma}} \right)^{\frac{1}{1-\frac{1}{\varphi}}}$$

s.t.

$$w' = (w - c)R^*(x, x'),$$

$$R^* = \vartheta R(x, x') + (1-\vartheta)R_f(x),$$

$$x' = lom(x),$$

where here we use x as a shorthand for both state variables, and w total wealth normalized by GDP, with a slight abuse of notation.

Exploiting the homotheticity in the Bellman equation, we can eliminate w and therefore reduce the dimensionality of the state space by one. We define $v(x)$ as $V(x, w) = v(x)w$. The Bellman equation is then given by:

$$v(x)^{1-\frac{1}{\varphi}} \frac{w^{1-\frac{1}{\varphi}}}{1-\frac{1}{\varphi}} = \max_{c, \theta} (1-\beta) \frac{c^{1-\frac{1}{\varphi}}}{1-\frac{1}{\varphi}} + \beta \frac{\mathbb{E} [v(x')^{1-\gamma} w'^{1-\gamma}]^{\frac{1-\frac{1}{\varphi}}{1-\gamma}}}{1-\frac{1}{\varphi}}$$

$$= \max_{c, \theta} (1-\beta) \frac{c^{1-\frac{1}{\varphi}}}{1-\frac{1}{\varphi}} + \beta \frac{(w-c)^{1-\frac{1}{\varphi}}}{1-\frac{1}{\varphi}} \mathbb{E} [v(x')^{1-\gamma} R^{*1-\gamma}]^{\frac{1-\frac{1}{\varphi}}{1-\gamma}}.$$

The maximization problem is solved in two steps. First, we take the first order condition with respect to θ :

$$\mathbb{E} [v(x')^{1-\gamma} (\theta^* R + (1-\theta^*)R_f)^{-\gamma} (R - R_f)] = 0,$$

which gives θ^* , independent of w . Defining $A(x) = \mathbb{E}[v(x')^{1-\gamma} R^{*1-\gamma}]^{\frac{1-\frac{1}{\varphi}}{1-\gamma}}$, then optimal consumption solves:

$$v(x)^{1-\frac{1}{\varphi}} \frac{w^{1-\frac{1}{\varphi}}}{1-\frac{1}{\varphi}} = \max_c (1-\beta) \frac{c^{1-\frac{1}{\varphi}}}{1-\frac{1}{\varphi}} + \beta A(x) \frac{(w-c)^{1-\frac{1}{\varphi}}}{1-\frac{1}{\varphi}}.$$

The first-order condition is given by:

$$(1-\beta)(c^*)^{-\frac{1}{\varphi}} = \beta A(x)(w-c^*)^{-\frac{1}{\varphi}},$$

which yields:

$$\frac{c^*(w)}{w} = \frac{1}{\left(\frac{\beta}{1-\beta}\right)^\varphi A^\varphi + 1} = \zeta(x).$$

Again, the consumption-wealth ratio is only a function of x but not w . Plugging it into the Bellman equation:

$$\begin{aligned} v(x)^{1-\frac{1}{\varphi}} &= (1-\beta)\zeta^{1-\frac{1}{\varphi}}(x) + \beta A(x)(1-\zeta(x))^{1-\frac{1}{\varphi}}, \\ v(x) &\equiv \mathbb{F}(v)(x). \end{aligned}$$

Since it is derived from a Bellman equation, the operator \mathbb{F} is also a contraction mapping. Therefore, we can solve it by iteration.

I.4.3 Calculating the elasticity with labor income

Now we proceed to calculate the macro elasticity, recognizing that only a fraction ψ of total wealth is capitalized.

Recall the definition of macro elasticity:

$$\zeta^r = -\frac{dQ^\varepsilon/Q^\varepsilon}{dP/P},$$

where $\psi = \frac{Q^\varepsilon}{Q}$. Therefore, the elasticity can be calculated as:

$$\zeta^r = -\frac{d\theta}{dp} \frac{Q}{Q^\varepsilon} = \frac{d\theta}{dp} \frac{1}{\psi}.$$

The share ψ is calibrated to match $\frac{Y}{W^\varepsilon} = \frac{Y}{\psi W} = \frac{1}{\psi(pd+1)} = 0.8$, and $\frac{d\theta}{dp}$ can be computed by perturbing pd as outlined above.

I.4.4 Calibration

The calibration closely follows Bansal and Yaron (2004), with two exceptions. In the original paper, Bansal and Yaron (2004) obtain a high risk premium from leveraged dividends. In our model, as we do not have an explicit dividend process, we calibrate $\bar{\sigma}$ to match a 4% risk premium on the aggregate market. We also reduce β so that we have a reasonable risk-free rate and P/D ratio in the stationary equilibrium.

References for Online Appendix

- 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.
- Barro, Robert J**, “Rare disasters and asset markets in the twentieth century,” *The Quarterly Journal of Economics*, 2006, 121 (3), 823–866.
- 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.
- Betermier, Sebastien, Laurent E Calvet, and Evan Jo**, “A supply and demand approach to equity pricing,” 2019.
- 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.
- 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.
- 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.
- 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.
- Evans, Martin DD and Richard K Lyons**, “Order flow and exchange rate dynamics,” *Journal of political economy*, 2002, 110 (1), 170–180.
- Frazzini, Andrea, Ronen Israel, and Tobias J Moskowitz**, “Trading costs,” *Working Paper*, 2018.

- 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 and Ralph SJ Koijen**, “Granular instrumental variables,” Working Paper 28204, National Bureau of Economic Research December 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.
- Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus**, “Five Facts About Beliefs and Portfolios,” *American Economic Review*, 2021, *111* (5), 1481–1522.
- Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus**, “The joint dynamics of investor beliefs and trading during the COVID-19 crash,” *Proceedings of the National Academy of Sciences*, 2021, *118* (4).
- Hasbrouck, Joel**, *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*, Oxford University Press, 2007.
- Johnson, Timothy C**, “Dynamic liquidity in endowment economies,” *Journal of Financial Economics*, 2006, *80* (3), 531–562.
- Kahneman, Daniel and Dan Lovallo**, “Timid choices and bold forecasts: A cognitive perspective on risk taking,” *Management science*, 1993, *39* (1), 17–31.
- 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.

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

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.