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

Xavier Gabaix and Ralph S.J. Koijen *

July 8, 2020. Preliminary : Comments most welcome.

Abstract

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

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

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

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

*xgabaix@fas.harvard.edu, ralph.koijen@chicagobooth.edu. We thank Antonio Coppola, Dong Ryeol Lee, Hae-Kang Lee, Simon Oh and Lingxuan Wu for excellent research assistance. We thank John Campbell, Robin Greenwood, Lars Hansen, Sam Hanson, John Heaton, Arvind Krishnamurthy, Alan Moreira, Toby Moskowitz, Stefan Nagel, Jonathan Parker, Lasse Pedersen, Andrei Shleifer, Jeremy Stein, Adi Sunderam, Motohiro Yogo, and seminar participants at AQR, Brandeis, Chicago, Cornell, Dartmouth, Harvard, IMF, virtual finance, virtual finance theory, virtual macroeconomics for comments. Gabaix thanks the Sloan Foundation for financial support. Koijen acknowledges financial support from the Center for Research in Security Prices at the University of Chicago Booth School of Business.

1 Introduction

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

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

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

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

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

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

The empirical core of this paper is to provide a quantification of the market's aggregate elasticity. To do that, we use a new instrumental variables approach, which was conceived for this paper and worked out in a stand-alone paper (Gabaix and Koijen (2020)), the "granular instrumental variables" (GIV) approach. The key idea is that we use the idiosyncratic demand shocks of large institutions or sectors as a source of exogenous variation. We extract these idiosyncratic shocks from factor models estimated on the changes in holdings of various institutions and sectors. We then take the size-weighted sum of these idiosyncratic shocks (the GIV), and use it as a primitive instrument to see how these demand shocks affect aggregate prices and the demand of other investors. This way, we can estimate both the aggregate sensitivity of equity prices to demand shocks (which is the multiplier of around 5 we mentioned above) and the demand elasticity of various institutions (around 0.2). Importantly, the data are consistent with a quite long-lasting price impact of flows. Indeed, in the simplest version of the model, the price impact is perfectly long-lasting. This is not because flows release information, but instead simply because of the permanent shift in the demand for stocks. We perform a large number of robustness checks, for example using different data sets (the Flow of Funds as well as 13F filings). The findings are consistent across specifications, in the sense that the price impact multiplier remains around 5 — indeed, in the range of 3.5 to 9. Furthermore, lots of things fall into place then: for instance, the volatility of the market can be traced back to the volatility of flows and demand changes.

We wish to highlight that the "inelastic market hypothesis" remains a hypothesis. Our empirical analysis relies on a new empirical methodology and on fairly unexplored data in this context. We hope that similar analyses will be performed, for instance in other markets, using other time periods, and using different methodologies. 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. Yet, despite the decades of research on investors' risk aversion, their elasticity of inter-temporal substitution, and the "micro" elasticity of demand for individual stocks, we have no estimates of the "macro" elasticity of demand for the equity market portfolio. We provide a first estimate, and we hope that future research will explore other identification strategies to improve and sharpen this estimate.

Our findings are not only at odds with traditional theoretical models, but also with the prevailing common wisdom in our profession at the time of writing this paper We quantify this via two surveys. We conducted a first survey by putting out a request via Twitter (using the #econtwitter tag) to complete an online survey. In addition, we asked participants of an online seminar at VirtualFinance.org to complete the same survey – this latter audience being naturally more representative of the population of academic researchers in finance. Both surveys were conducted before the paper was available online and before the seminar was conducted. We received 197 responses for the Twitter survey and 105 responses for the survey connected to the finance seminar.

The survey question was the following: "If a fund buys $1 billion worth of US equities (permanently; it sells bonds to finance that position), slowly over a quarter, how much does the aggregate market value of equities change?" The answer given in this paper is M times a billion, where M is the price impact multiplier, which we estimate to be around $M = 5$. At the same time, the simplest "efficient markets" answer is $M = 0$: in that view, the price is the present value of future dividends, unperturbed by flows.[1] In both surveys, the median answer was $M = 0$: surveyed economists, logically enough, rely on the basic finance model with efficient markets. In both samples, the median answer, conditionally on being strictly positive, was $M = 0.01$.[2] Hence, *surveyed economists' priors underestimate the price impact of flows by a factor of over 500.*

[1] In fact, the rational M is a bit higher than that, as we detail in Section 3.5.

[2] The answer $M \geq 1$ was given by only 2.5% of respondents in the Twitter survey and by 4% of respondents in the VirtualFinance.org survey.

Here are four a priori reasons to entertain that markets would be inelastic First of all, if one wants to buy $1 worth of equities, many funds actually cannot supply that: for instance, a fund that invests entirely in equities cannot exchange them for bonds. Many institutions have tight mandates, something that we confirm in this paper. Second, it is hard to find the "would-be arbitrageurs." For instance, hedge funds are small (they hold only about 5% of the equity market), and they tend to reduce their equity allocations in bad times (for example, because of outflows or risk constraints; see Ben-David et al. (2012)). Third, the transfer of equity risk across investor sectors is small (about 0.6% of the aggregate value of the equity market per quarter for the average pair of investor sectors). This suggests again that the demand elasticity of most investors is quite small. Fourth, a substantial literature has shown that when a company is removed from an index, its share price falls, where the latest estimates find a "micro" demand elasticity of approximately one (we give complete references in the literature review below). As the macro elasticity should arguably be lower than the micro elasticity (considering that, for example, Apple and Amazon are closer substitutes than the stock market index and a bond), this suggests a low macro elasticity, perhaps less than one.

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

If equity markets are indeed inelastic, several questions that are irrelevant or uninteresting in traditional models become interesting. For instance, government interventions matter. We discuss potential policy implications of our approach. As an example, in our model, 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.

Also, firms as financiers materially impact the market. Prior research showed that firms react to price signals, such as in their decisions to issue dividends or raise funds in stocks versus bonds (Baker and Wurgler (2004); Ma (2019)): now we can quantity how firms' actions impact the market. For instance, stock buybacks can have a large aggregate effect. Suppose that the corporate sector buys back $1 worth of equities. In the traditional Modigliani-Miller world, the value of the remaining shares does not change at all. In contrast, in an inelastic world, the value of the remaining shares goes up, by a tentative estimate of between $0.5 and $2.[3] As a naive non-economist might think, "if firms buy shares, that drives up the price of shares" — but here the naive thinking is actually qualitatively correct. Hence, potentially, as share buybacks account for a large portion of flows (they have been about as large as dividend payments in the recent decade), corporate actions account for a sizable share of equity purchases, and hence of the volatility and valuation growth of the stock market. This "corporate finance of inelastic markets" is an interesting avenue of research.

We then show how to construct a general equilibrium model in the spirit of Lucas (1978), but

[3]For that estimate, we need to conjecture the reaction of consumers to share buybacks, based on some existing studies.

where there is a central role for flows and inelasticity. It clarifies the role of flows, the determination of the interest rate, and shows how to augment traditional general equilibrium models with flows in inelastic markets. That makes those models more realistic, and better suited for policy. This model may serve as a prototype for models enriched by inelasticity. We also show how to marry flows to the "stochastic discount factor" (SDF) approach: the flows are primitive, and the SDF is a book-keeping device to record their influence on prices. This model could be helpful to get correct risk prices in macroeconomic models, including their variation due to flows.

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

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.

A literature studies the impact of mutual fund flows in the market, including Warther (1995) and Frazzini and Lamont (2008). One innovation of this paper is to provide a framework to think about this, to include all sectors (not just mutual funds), and think about causal inference via the GIV.

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

A related literature finds convincing evidence that supply and demand changes do affect prices and premia in partially segmented markets, for bonds (for example as in Greenwood and Vayanos (2014), Greenwood and Hanson (2013), and Vayanos and Vila (2020)), mortgage-backed securities (Gabaix et al. (2007)), or options (Garleanu et al. (2009)), with models which typically feature CARA investors. Here our focus is on stocks, and our model is quite different from the models in that literature.

We also relate to the literature on slow-moving capital (Mitchell et al. (2007); Duffie (2010); 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 microstructure literature (Kyle (1985)). There, inflows cause price changes, but crucially those inflows do not change the risk premium on average (as the mechanism is rational Bayesian updating, rather than limited risk-bearing capacity), and hence do not create excess volatility. In contrast, in our paper, inflows do change the risk premium, creating excessively volatile prices.

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

Notations For a vector $X = (X_i)_{i=1\ldots N}$ and a series of relative weights S_i with $\sum_{i=1}^N S_i = 1$, we let

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

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

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

We use $\mathcal{E}$ for equities, $\mathbb{E}$ for expectations, and E for equal-weighted averages.

We generally use lowercase notations for deviation from a baseline. We call δ the average dividend-price ratio of the equity market.

2 Initial Motivations to Explore the Inelastic Markets Hypothesis

The central tenet of the inelastic markets hypothesis is that the aggregate stock market is inelastic. In this section, we provide three pieces of a priori evidence that the inelastic market hypothesis is plausible. These facts are meant to be no more than suggestive: the core empirical results are in Section 4, in which we provide a careful quantification of the key parameters of our model.

We first seek to identify investors with elastic demand for the aggregate stock market in Section 2.2. That is, we ask: who are the deep-pocketed arbitrageurs that could make the aggregate stock market elastic? This question relates to the work by Brunnermeier and Nagel (2004), who show that hedge funds did not provide elasticity to the market during the technology bubble in the late nineties. Next, we show that the flows between sectors are relatively small and that small changes in quantities are correlated with large changes in prices, a sign of inelastic markets. We conclude by connecting the inelastic markets hypothesis to the evidence on micro elasticities, which also suggests that the this hypothesis may hold. We use data described in Section 4.1, primarily here the Flow of Funds (FoF).

2.1 Who owns equities and how do their equity shares fluctuate over time?

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

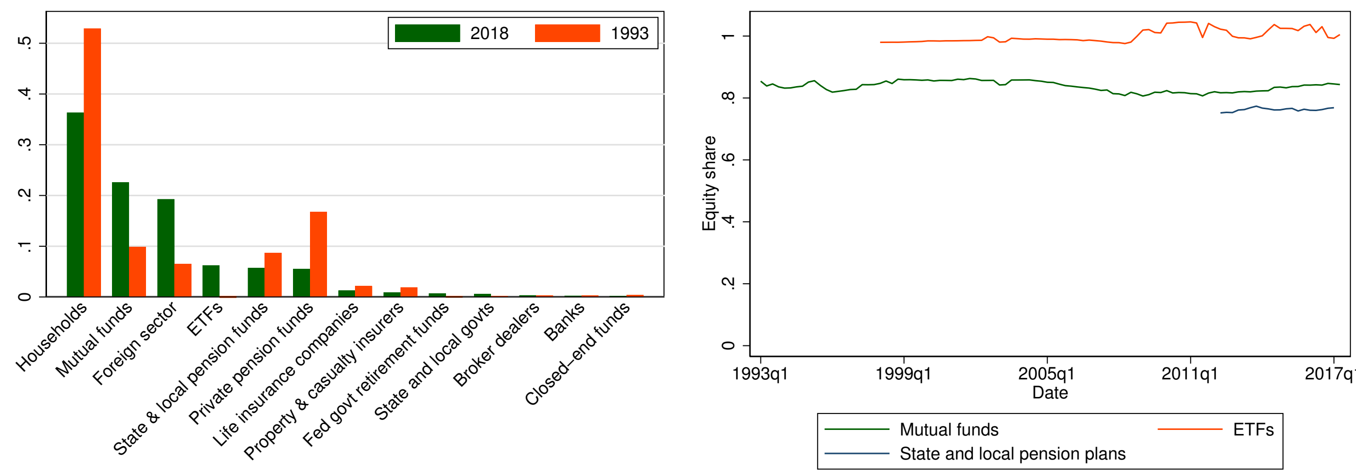
For some of these sectors, such as mutual funds, exchange-traded funds, and pension funds, we have investor-level data on equities and fixed income holdings, albeit for different sample periods. In the right panel of Figure 1, we plot the equity share. Consistent with the theory that we develop in the next section, we aggregate different investors in a given sector using the relative size of their equity portfolios (as opposed to assets under management). The main takeaway is that the equity shares are stable over time.[4]

2.2 In search of the arbitrageurs: Who rebalances during equity downturns?

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

[4]For mutual funds, funds can be broadly categorized into three groups, equity funds, bond funds, and mixed funds. The first two categories provide little to no elasticity to the market. The mixed funds, such as target-date funds, oftentimes follow close to fixed share strategies. Combining these groups lead to the stable pattern observed in Figure 1. For ETFs, the category of mixed funds is smaller but the same intuition applies.

Figure 1: Equity shares. The left panel of the figure plots the equity share in 1993 (orange bars) and in 2018 (green bars) by institutional sector. The right panel displays the value-weighted average equity share of mutual funds, ETFs, and state and local pension plans. The equity share of the different institutions are averaged using the relative equity size of each investor. The construction of the data are discussed in Appendix Section 9.



(the technology crash) and from 2007.Q4 to 2009.Q1 (the 2008 global financial crisis), as shown in Appendix Section 10.1. To measure equity flows, we scale the dollar equity flows, $\Delta F_{jt}^{\mathcal{E}}$, 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 then average the percent flow by sector across quarters for a given downturn.

The left panel of Figure 2 corresponds to the tech crash and the right panel to the 2008 financial crisis. In the case of the 2008 financial crisis, we separately report the results for 2008.Q4, which is the worst quarterly return in our sample. In both cases, we select the eight sectors with the largest absolute flows as well as the corporate sector. While the total equity risk reallocation, on average per quarter, remains small, households sell about 0.5% of the market per quarter.[5]

During the 2008 financial crisis, net repurchases by firms turned negative, implying that they issued equity. If we zoom in on 2008.Q4, we see large issuances (for instance by financial firms),[6] 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 simply reflect their mandate to maintain a fixed-share strategy instead of a conscious effort to time the market (see Proposition 2).[7]

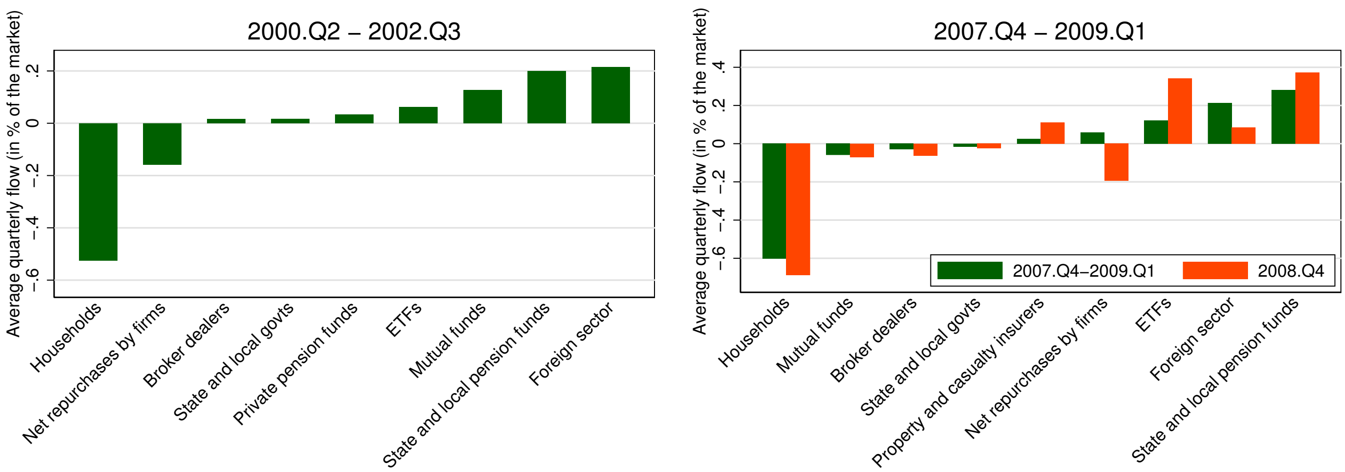
Hedge funds, as a whole, are small, and reduce equity holdings in downturns. We conclude by zooming in on one particular group of intermediaries that one can view as the prime

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

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

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

Figure 2: The figure illustrates the rebalancing of investors during drawdowns of the U.S. 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.



example of smart arbitrageurs, namely hedge funds.[8]

First, hedge funds own a small fraction of all US equities: they hold around 5% of the U.S. stock market in 2016. Second, hedge funds reduce their equity holdings during equity downturns, see Ben-David et al. (2012), which is the opposite of what we would expect to find if hedge funds are particularly macro price-elastic. Economically, their trading behavior may simply reflect that hedge fund investors withdraw their capital or that risk constraints bind,[9] which forces hedge funds to reduce their leverage and overall riskiness of their portfolios. Still, the combination of their own risk constraints and the behavior of their clients leads hedge funds to be small, and to reduce equity holdings in downturns. Hence, hedge funds do not provide appreciable elasticity to markets in downturns.

2.3 Flows across investor classes are small

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

[8]While a large literature explores the micro elasticity of hedge funds, we are interested in their market elasticity. In the FoF, hedge funds are part of the household sector and we cannot study them separately using these data. 13F data allows one to do this accurately though.

[9]See Adrian and Shin (2013) and Coimbra and Rey (2020) for models of intermediaries featuring value-at-risk constraints.

repurchases, we separately measure the equity risk "creation" and "redemption" as a result of such corporate actions via $y_t^{AbsNet} = \frac{|\Delta F_t^{Firm}|}{\mathcal{E}_{t-1}}$, which we will refer to as absolute net flows.

The average absolute net flow equals 0.30% per quarter and the average gross flows average to 0.87% per quarter for the period from 1993.Q1 to 2018.Q4. The standard deviations are 0.26% and 0.37%, respectively. The difference between the series measures the risk reallocation in equity markets across institutional sectors, which averages to approximately 0.6% per quarter.[10] We plot the time series of both measures in Figure 12 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 Section 10.3 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.

2.4 A bound based on the micro elasticity of demand

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

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

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

3 Basic Theory with Inelastic Markets

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

3.1 Two-period model for the aggregate stock market

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

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

We consider that the mixed fund's demand for stocks Q^D is given by a mandate, saying that the fraction of the portfolio invested in equities should be equal to the following:

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

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

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

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

Prices react a lot to flows in inelastic markets Initially (in a baseline pre-period $t = 0^-$, before any shocks) the fund has wealth $\bar{W}$, and holds all the equity shares. We assume that before the shocks, equities have a risk premium $\bar{\pi}$ and the dividend-price ratio is at its corresponding value, $\delta = \frac{\bar{D}^e}{\bar{P}}$, where $\bar{P}$, $\bar{D}^e$ are the baseline values for the stock's price and the expected dividend. Then, at time 0, suppose that the representative consumer decides (for whatever reason) to sell ΔF dollars of the pure bond fund, and invest them in the mixed fund. What happens then? The answer is given in the next proposition. It is exact for $\kappa = 0$, and uses a first-order Taylor expansion for small flows f when $\kappa \neq 0$.

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

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

where ζ *is the macro elasticity of demand:*

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

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

We shall soon calibrate $\zeta \simeq 0.2$, so that $\frac{1}{\zeta} \simeq 5$. Then, (4) means that if investors buy 1% of the equity market (selling bonds), the price of equities increases by 5%. This is symmetric: if investors sell 1% of the market, the price of equities falls by 5%. One can also say it in dollar terms. If someone buys $1 worth of equities (selling $1 worth of bonds), the market value of equities increases by $5.

We next derive this result.

Deriving the elasticity of the demand for stocks We first derive the demand for stocks. We do that under a slightly more general assumption, assuming that at time 0 there may be a change d in the value of fundamentals.

We call p, d, q^D the fractional deviations of the price, expected dividend, and demand from their baseline values:

$$p = \frac{\Delta P}{\bar{P}} = \frac{P}{\bar{P}} - 1, \qquad d = \frac{D^e}{\bar{D}^e} - 1, \qquad q^D = \frac{Q^D}{\bar{Q}^D} - 1, \qquad f = \frac{\Delta F}{\bar{W}}. \tag{6}$$

We perform the analysis for small disturbances f, d, and hence small p, q^D, here and throughout the paper.[11]

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

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

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

Before deriving this Proposition 2, we observe that it delivers Proposition 1 as a simple consequence. We assume that the supply of shares does not change (we shall relax this later). So, in equilibrium we must have $q^D = 0$. This and (7), with $d = 0$, give the price, $p = \frac{f}{\zeta}$. We next prove Proposition 2.

Proof. We observe that before the shock, we had

$$\bar{W} = \bar{P}Q + \bar{B},$$

where $\bar{P}Q$ and $\bar{B}$ are respectively the fund's holdings of equities and bonds:

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

[11] Following a common convention in macro-finance, we do Taylor expansions of the leading term, omitting the formal mentions of $O(\cdot)$ terms.

After the shock,[12] we have $W = PQ + \bar{B} + \Delta F$ as the price has changed to P, and the fund's value is made of the equity value PQ (at the new price P) and the inflow ΔF, so that

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

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

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

that is

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

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

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

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

so

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

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

Next, consider the case with a general κ*. (We recommend skipping this part at the first reading.)* We note that the price-dividend ratio is $\delta = \frac{D^e}{P} = 1 + r + \pi$. Taking logs and then deviations from the baseline, $\frac{\hat{\pi}}{1+r+\pi} = d - p$, that is

$$\hat{\pi} = \delta\,(d-p)\,, \tag{10}$$

so that using (9),[13]

$$q^D = w - p + \kappa\hat{\pi}, \tag{11}$$

that is

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

which yields (7). □

We now comment on the economics of this Proposition 3.

[12] At this stage, when $\kappa = 0$, there is a simpler proof, which however does not show that ζ is the elasticity of demand. We have $\bar{P}Q = \frac{\theta}{1-\theta}\bar{B}$ before the shock, and $\left(\bar{P} + \Delta P\right) Q = \frac{\theta}{1-\theta}\left(\bar{B} + \Delta F\right)$ after, as the bond holding of the mixed fund must increase by ΔF after the inflow: the mixed fund cannot sell stocks to any other outside entity. Subtracting those two equations gives:

$$p := \frac{\Delta P}{\bar{P}} = \frac{\frac{\theta}{1-\theta}\frac{\Delta F}{Q}}{\bar{P}} = \frac{1}{1-\theta}\frac{\Delta F}{\frac{\bar{P}Q}{\theta}} = \frac{1}{1-\theta}\frac{\Delta F}{\bar{W}} = \frac{f}{1-\theta} = \frac{f}{\zeta}.$$

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

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

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

An undergraduate example We propose a simple, undergraduate-level example. Suppose that the representative mixed fund always holds 80% in equities ($\theta = 0.8, \kappa = 0$), so that $\zeta = 0.2$ and $\frac{1}{\zeta} = 5$. Then *an extra \$1 invested in the stock market increases the total market valuation by \$5.* This great ability of the market to multiply shocks will play a central role, and indeed will be a source of the "excess volatility" of stocks.

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 (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 \$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.[14,15]

Inflows not mediated by the fund have the same price impact $\frac{1}{\zeta}$ This effect is quite general. Suppose that there is a flow f^D into stocks, in the sense that a segment of the market inelastically buys a fraction f^D of the shares, and that there is no flow into the mixed fund nor news about dividends ($f = d = 0$). As in equilibrium the total demand change must be zero, the equilibrium is $f^D + q^D = 0$ with $q^D = -\zeta p$ (by equation (7)). That yields

$$p = \frac{f^D}{\zeta}.$$

So, if f^D percent of the shares are bought inelastically, the price increases by $\frac{f^D}{\zeta}$ percent.

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

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

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

[15] The same holds if bond prices are elastic (e.g. are a composite of long term bonds with elastic price, and short term bonds with inelastic price). Then, after the \$1 flow, bond prices do not change, but stock prices do change as in (4). This is because the total demand for bonds has not changed.

$$p=\frac{f}{\zeta}+M^{D}d, \tag{12}$$

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

$$M^{D}=\frac{\kappa\delta}{\zeta}=\frac{\kappa\delta}{1-\theta+\kappa\delta}\in[0,1]\,. \tag{13}$$

Proof. We assume that the supply of shares does not change (we shall relax this later). So, in equilibrium we must have $q^D=0$. This and (7) give the price, $p=\frac{f}{\zeta}+\frac{\kappa\delta}{\zeta}d$. □

Equation (12) entails a partial myopia, which is a second and more minor force in inelastic markets. If there is a shock to expected fundamentals, $\frac{\Delta D^e}{D^e}=d$, then Proposition 3 says that the price impact is

$$p=M^{D}d,$$

with $M^D\in[0,1]$ given in (13).[16] This implies that prices move less than one-for-one with expected fundamentals, as long as there is no flow f from the households correcting this. This is the flip side of inelasticity. The market is very forward-looking ($M^D\to 1$) only when $\kappa\to\infty$ (so that investors buy a lot of stocks when the risk premium is a bit higher than usual), or $\theta\to 1$, and $M^D=0$ when the market doesn't react to fundamentals (which happens for example if traders do not react to the risk premium, $\kappa\to 0$).

3.2 Infinite horizon model for the aggregate stock market

We extend the static model to a dynamic one. The forces will generalize, in an empirically implementable way. There is again a constant risk-free rate r, taken here to be exogenous (Section 6 endogenizes it in general equilibrium, but here we concentrate on the core economics of inelasticity). The representative stock gives a dividend D_t.

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

$$\frac{P_tQ_t^D}{W_t}=\theta e^{\kappa\hat{\pi}_t+\nu_t}, \tag{14}$$

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

We define by $\bar{P}_t$, $\bar{W}_t$, and $\bar{D}_t$ the baseline price, wealth, and dividend — the values before any flow or extra news friction.[17] We call p_t, w_t, d_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 likewise for $q_t^D=\frac{Q_t}{\bar{Q}_t^D}-1$. We define the flow f_t as the *sum* of past fractional inflows, since a date $t=0$ when the market was in equilibrium ($\hat{\pi}_0=0$):

$$f_t=\sum_{s=0}^{t}\frac{\Delta F_s}{\bar{W}_{s-1}}, \tag{15}$$

[16] We assume $\theta\in(0,1)$. In leveraged markets, we'd have $\theta>1$, and $M^D>1$.

[17] For simplicity, we postpone the formal definition of those "baseline" values, but they are exactly defined in the context of the economy of Proposition 10 and its proof.

where ΔF_s are dollar inflows.[18] In this definition, we assume that dividends are passed to consumers, which implies that reinvested dividends count as new flows. We call again δ the typical dividend-price ratio, $\delta := \overline{D/P} = r + \bar{\pi} - g$, with g the growth rate of dividends. We call the expected dividend deviation $d_t^e = \mathbb{E}_t d_{t+1}$. The aggregate demand for stocks is as follows, generalizing (7).

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

$$q_t^D = -\zeta p_t + f_t + \nu_t + \kappa \delta d_t^e + \kappa \mathbb{E}_t \left[\Delta p_{t+1}\right], \tag{16}$$

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

Proof. The expected excess return is $\pi_t = \frac{\mathbb{E}_t[\Delta P_{t+1} + D_{t+1}]}{P_t} - r$, and we use the Taylor expansion:

$$\hat{\pi}_t = \mathbb{E}_t \left[\Delta p_{t+1}\right] + \delta \left(d_t^e - p_t\right). \tag{17}$$

The wealth deviation is, using the same derivation as in the two-period model, $w_t = \theta p_t + f_t$, so that the demand response is given by the linearization of (14):

$$\begin{aligned} q_t^D &= w_t - p_t + (\kappa \hat{\pi}_t + \nu_t) \\ &= (f_t + \theta p_t) - p_t + (\kappa \hat{\pi}_t + \nu_t) \qquad (18) \\ &= f_t + \nu_t - (1-\theta) p_t + \kappa \left(\mathbb{E}_t \Delta p_{t+1} + \delta \left(d_t^e - p_t\right)\right), \qquad (19) \end{aligned}$$

which yields (16). □

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

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

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

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

$$\rho = \frac{\zeta}{\kappa} = \delta + \frac{1-\theta}{\kappa}, \tag{21}$$

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

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

[18] This notation is very close to another natural one (replacing $\bar{W}_s$ by W_s in (15), where W_s is the actual fund holdings) but it leads to cleaner expressions over long time horizons.

Proof. Equation (16) can be rewritten as $q^D = \kappa\left(\mathbb{E}_t\Delta p_{t+1} - \rho p_t + \delta d_t^e\right) + f_t + \nu_t$. As $q^D = 0$, this is also:

$$\mathbb{E}_t\Delta p_{t+1} - \rho p_t + \delta d_t^e + \frac{\nu_t + f_t}{\kappa} = 0. \tag{23}$$

Defining $z_t := \delta d_t^e + \frac{\nu_t + f_t}{\kappa}$, this gives $p_t = \frac{\mathbb{E}_t p_{t+1} + z_t}{1+\rho}$, so that $p_t = \mathbb{E}_t \sum_{\tau \geq t}^{\infty} \frac{z_\tau}{(1+\rho)^{\tau - t + 1}}$. The risk premium comes from (18). □

In (20), the price discounts future dividends at a rate ρ given in (21).[19,20] So, *the market is more myopic (higher ρ) when it is less sensitive to the risk premium (lower κ) and when the mixed fund has a lower equity share (lower θ).*[21] It makes good sense that a lower sensitivity to the risk premium makes the market less reactive to the future, hence more myopic.[22,23] In the rest of this section, we set $\nu_t = 0$; otherwise, we observe that the general case comes from replacing f_t by $f_t + \nu_t$.

The impact of a flow that does not mean-revert Suppose that at time 0 there is an inflow f_0 that does not mean-revert. Then, the impact on the time$-t$ price is (via (20), with $\mathbb{E}_0\left[f_\tau\right] = f_0$, or (23)):

$$\mathbb{E}_0\left[p_t\right] = \frac{1}{\zeta} f_0. \tag{24}$$

So, the "price impact" is permanent. As a result, the equity premium is permanently lower, $\hat{\pi} = -\delta \frac{f_0}{\zeta}$. Again, this is not due to any informational channel.[24] This is simply because, if the equity demand has permanently increased, equity prices should be permanently higher.

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 $\phi \in [0, 1]$, so that the cumulative flow is $\mathbb{E}_0\left[f_\tau\right] = (1-\phi)^\tau f_0$. Then, the impact on the time$-t$ price is (via (20)) $p_t = \frac{1}{\zeta + \kappa\phi} f_t$, implying

$$\mathbb{E}_0\left[p_t\right] = \frac{(1-\phi)^t}{\zeta + \kappa\phi} f_0. \tag{25}$$

So, the "price impact" is smaller for flows that mean-revert more quickly: it is $\frac{1}{\zeta + \kappa\phi}$. Also, it mean-reverts at the rate ϕ, which is exactly the rate of mean-reversion of the cumulative flow.

---

[19] Indeed, a permanent dividend increase of d_t, not accompanied by flows, leads to a price increase of $p_t = M^D d_t$, like in (13).

[20] This myopia in (20) generates momentum: because the market is myopic (by (20)), dividend news are only slowly incorporated into the price.

[21] The formula extends to changes in the interest rate, as in $r_{ft} = \bar{r}_f + \hat{r}_{ft}$. As (17) becomes $\hat{\pi}_t = \mathbb{E}_t\Delta p_{t+1} + \delta\left(d_t^e - p_t\right) - \hat{r}_{ft}$, all expressions are the same, replacing d_t^e by $d_t^e - \frac{1}{\delta}\hat{r}_{ft}$, including in (20). In particular, the impact of a permanent change in the interest is muted by the same impact M^D.

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

[23] Here the demand (14) depends on the risk premium as $\kappa\hat{\pi}_t = \kappa\mathbb{E}_t\Delta p_{t+1} + \kappa\delta\left(d_t^e - p_t\right)$. A variant would be 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 $\kappa\mathbb{E}_t\Delta p_{t+1} + \kappa^D\delta\left(d_t^e - p_t\right)$ 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, with $M^D = \frac{\kappa^D\delta}{1-\theta+\kappa^D\delta}$. This highlights that κ^D increases market elasticity ζ, while κ increases market "forward-lookingness" $\frac{1}{\rho}$.

[24] In a Kyle (1985) model, flows change prices, like in our model; but they do not on change the risk premium (on average), which is a crucial difference with our model. Section 12.8 details the link with the Kyle model.

A simple benchmark It is illustrative to consider a simple benchmark:

$$f_t = (1-\theta)\, d_t + \tilde{f}_t, \tag{26}$$

where d_t is the past cumulative dividend growth, taken to be a martingale, and $\tilde{f}_t$ is autoregressive with speed of mean-reversion ϕ_f. Then, (20) gives:

$$p_t = b_f^p \tilde{f}_t + d_t, \qquad \hat{\pi}_t = b_f^\pi \tilde{f}_t, \qquad b_f^p = \frac{\rho}{\rho+\phi_f}\frac{1}{\zeta} = \frac{1}{\zeta+\kappa\phi_f}, \qquad b_f^\pi = -\left(\delta+\phi_f\right) b_f^p. \tag{27}$$

A high inflow increases equity prices and hence lowers the risk premium.[25]

3.3 Heterogeneous funds

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

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

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

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

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

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

$$\theta = \theta_S, \qquad \kappa = \kappa_S, \qquad \zeta = \zeta_S, \qquad f = f_S. \tag{28}$$

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

The equity-weighted equity share always exceeds the wealth-weighted equity share, $\theta_S \geq \mathbb{E}_W[\theta_i]$.[26] The former is what matters for the macro elasticity, while the latter is directly available in aggregated data. This makes the disaggregation issues potentially non-trivial, and will require some care in the empirical part.

[25] Relatedly, call p_t^* the "rational" price that keeps a constant risk premium ($p_t^* = \sum_{\tau=t}^{\infty} \frac{\delta}{(1+\delta)^{\tau-t+1}} d_\tau^e$). Then, the "rational" flow that ensures a constant risk premium is simply $f_t^* = (1-\theta)\, p_t^*$.

[26] Indeed, using $\mathcal{E}_i = \theta_i W_i$ and (in the last step) the Cauchy-Schwarz inequality:

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

Infinite-horizon model Fund i has demand (16), $q_{it} = -\zeta_i p_t + f_{it} + \nu_{it} + \kappa_i \delta d_t^e + \kappa_i \mathbb{E}_t [\Delta p_{t+1}]$. Aggregating over all funds i, we have (taking again the equity-weighted averages):

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

This means that (16) holds, replacing θ, κ, f_t, and ν_t by their the equity-weighted averages as in (28). As a result, the rest of the results in that section also hold (for example including (20)), using equity-share weighted averages.

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

$$f^F = \frac{\text{Net repurchases (in value)}}{\text{Total equity value}} = -\frac{\text{Net issuances (in value)}}{\text{Total equity value}}. \tag{29}$$

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

$$f := f_S + f^F, \tag{30}$$

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 consumers into equities, we want to add share repurchases by corporations.

3.4 A calibration

We provide a tentative calibration for the aggregate equity market, using yearly units throughout. We take a dividend-price ratio $\delta = \frac{D}{P} = 4\%/\text{year}$. Given the results in Figure 1, we take an equity share $\theta = 0.86$ (equity-weighted in the sense of Section 3.3). We also take $\kappa = 1.4$ years for the sensitivity of the equity share to the risk premium.[27] This is somewhat arbitrary, but helps calibrate the empirical value of ζ.

Given this, the macro elasticity is:

$$\zeta = 1 - \theta + \kappa\delta = 0.2, \tag{31}$$

and the price impact of flows is $\frac{1}{\zeta} = 5$.[28]

The market discount factor (21) is then:

$$\rho = \frac{\zeta}{\kappa} = 13\%/\text{year}. \tag{32}$$

[27] One motivation is that if the stock premium increases from $\pi = 5\%$ to $\pi = 9\%$, a flexible fund might change its equity allocation from 0.5 to 0.6, multiplying it by 1.2. So the flexible fund has $\kappa_i = \frac{d \ln \theta_i}{d\pi} = \frac{0.2}{0.04} = 5$. Most funds, however, are quite rigid: for instance 100% equity funds need to have $\kappa = 0$. So, we might tentatively parametrize $\kappa_i = H(1 - \theta_i)$, with $H \simeq 10$. So, we obtain $\kappa = \kappa_S = H(1 - \theta_S) = 10 \times 0.14 = 1.4$.

[28] Relying on Section 3.3, we could imagine that only a fraction m_p of funds are "active", while the rest are simply buy-and-hold, and hence provide no elasticity to the market. Then, the basic formula becomes (by the analysis in Section 3.3): $\zeta = m_p (1 - \theta + \kappa_a \delta)$, where κ_a is the forward-lookingness κ of active funds. We shall not use that degree of freedom of m_p in the calibration, but perhaps in some other countries or markets it might be useful.

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

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

First, as a partial intuition, if agents were risk neutral and the risk 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 "undergraduate" efficient markets hypothesis, where the price is always equal to the present value of dividends (with a constant discount rate), independently of flows.

Second, let us examine the more sophisticated case with risk-averse agents. We model aggregate income Y_t as going to the equity dividend as $D_t = \psi Y_t$, and the rest going to labor and other forms of business as $D_t^L = (1-\psi)\, Y_t$.

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. How much capital should flow into equities? Johnson (2006) analyzes a similar question with different assumptions and definitions, finding quite high macro elasticities in rational model.[29]

For simplicity, we consider the most classic case, with a CRRA consumer with utility $\sum_t e^{-\rho t} \frac{C_t^{1-\gamma}}{1-\gamma}$ and with the endowment Y_t following an i.i.d. growth process: $Y_t = G_t Y_{t-1}$. 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 $1 - p\Delta t$), and $G_t = e^{g\Delta t} B_t$ if there is a disaster (which happens with probability $p\Delta t$), where $B_t > 0$ is the (potentially stochastic) recovery rate of the economy. The proof of the next proposition is in Section 11.

Proposition 6. (Market elasticity in frictionless rational or behavioral models) *We derive the Marshallian price-elasticity of the demand for stocks in two classes of frictionless models. We suppose that all agents are frictionless (and with common beliefs, rational or behavioral), with CRRA utility and with i.i.d. endowment growth.*

Consider first the basic model in which growth rates are lognormal. Then, the elasticity of demand for equities is:

$$\zeta^r = \frac{1}{\pi} \frac{C}{W^{\mathcal{E}}}, \tag{33}$$

where π is the risk premium, C is consumption and $W^{\mathcal{E}}$ is the stock market capitalization.

Next, consider a disaster model where the growth rate follows a jump process. Then the elasticity of demand for equities is

$$\zeta^{r,D} = \frac{1}{\pi} \frac{C}{W^{\mathcal{E}}} \frac{B}{\gamma (1-B)}, \tag{34}$$

[29] Johnson (2006) also uses a different definition, which leads to $\zeta^r = \frac{\delta}{\gamma(g-\pi)}$ (so, it can be negative), where g is the growth rate of the economy, and π is the risk premium. In contrast, we find $\zeta^r = \frac{\delta}{\psi\pi}$ (a rewriting of (33)). An important difference is that we derive analytical expression assuming that only a fraction $\psi \in (0,1]$ of the tree is securitized. That is realistic (the aggregate stock market payout is indeed a very small part of GDP), and multiplies the elasticity by $\frac{1}{\psi}$. A more minor difference is that ours is the Marshallian demand, keeping the interest rate fixed, whereas his demand is Hicksian in spirit, and allows for a variable interest rate.

where B is the recovery rate of the endowment after a disaster.

Take the calibrated values $C = 0.8Y$, where Y is GDP, $W^{\mathcal{E}} = Y$ (as the typical market capitalization is roughly equal to GDP), and $\pi = 4\%$. Then (33) implies that the elasticity predicted by rational models is $\zeta^r = 20$. Hence, with an empirical elasticity $\zeta = 0.2$, we find that the basic rational model predicts an elasticity of demand 100 times bigger than the empirical one:

$$\frac{\zeta^r}{\zeta} = 100. \tag{35}$$

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 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 risk premium π in (33) by the perceived risk 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 (34) is $\zeta^{r,D} = 10$, so that it is 50 times larger than the empirical one. We strongly suspect that a similar reasoning would work for a habit formation model (Campbell and Cochrane (1999)) and a long run risks model (Bansal and Yaron (2004)).[30,31]

Why is the aggregate demand for equities so inelastic? The core of the inelastic markets hypothesis is that the macro demand elasticity ζ is low. Why is it so low?

We highlighted two reasons, namely fixed-share mandates, such as those of many funds that are 100% in equities and hence have zero elasticity (and in general $\zeta > 0$, $\kappa = 0$), 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, κ are negative. A different possibility is that when prices move, people's subjective perception of the risk 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 risk premium is a statistically error-prone procedure, so that market participants shrink towards no reaction to this. 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 P/E ratios are not that useful as short-run forecasters (Gabaix (2014), Chinco and Fos (2019)), or because many investors just do not wish to bother paying attention to them. The pass-through between subjective beliefs and actions might be low, as it is for retail

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

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

investors (Giglio et al. (2019)). Finally, demand may respond little to prices because demand shocks are highly persistent.[32]

In the end, while pinpointing the exact reasons for low market elasticity would be quite interesting, this question has a large number of plausible answers. Fortunately, it is possible to write a framework in a way that is relatively independent to the exact source of low elasticity, and this is the path we choose so far.

3.6 Micro versus macro elasticity: The cross-section of stocks

We generalize to a model with several stocks. This allows us to distinguish between the macro elasticity of demand for stocks, ζ and the micro-elasticity $\zeta^{\perp}$. The upshot is that the effects are the same, but with higher demand elasticity in the cross-section $\zeta^{\perp} > \zeta$ than in the aggregate. We recommend skipping this section at the first reading.

3.6.1 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 risk premium of asset a from the CAPM benchmark (this could be generalized of course), and $\hat{\pi}_{at}^{\perp} = \pi_{at}^{\perp} - \bar{\pi}_a^{\perp}$ as its deviation from the average.

We start from a model of stock-level demand for stock a (as in *a*sset), which comes from a "tracking error" type of mandate: the fraction in equities allocated to asset a is

$$\frac{P_{at} Q_{at}^D}{P_t Q_t^D} = \theta_a^{\mathcal{E}} e^{\kappa^{\perp} \hat{\pi}_{at}^{\perp} + \nu_{a,t}^{\perp} + \theta^{\perp} p_a^{\perp}}. \tag{36}$$

Indeed, $P_{at} Q_{at}^D$ is the dollar demand for asset a, and $P_t Q_t^D$ is the dollar demand for the aggregate stock market. On average, their ratio is $\theta_a^{\mathcal{E}}$. The term $\kappa^{\perp}$ is the micro-elasticity of demand with respect to the anomalous part of the risk premium $\hat{\pi}_{at}^{\perp}$. The term $\nu_{at}^{\perp}$ is an extra demand shock, with $\sum_a Q_a P_a \nu_{at}^{\perp} = 0$. The term $\theta^{\perp}$ indicates a concern for tracking error: if the fraction allocated to asset a is constant, then $\theta^{\perp} = 0$ (this is the baseline case). However, if the number of shares allocated to asset a is constant, then $\theta^{\perp} = 1$.

Calling $q_{at}^d = \frac{Q_{at}}{Q_a} - 1$ the deviation of the demand from the baseline, and $q_{at}^{D,\perp} = q_{at}^D - q_t^D$ how much asset a deviates from the baseline, we obtain the following counterpart to Proposition 4.

Proposition 7. *(Demand for individual stocks in the infinite-horizon model) The demand change (compared to the baseline) for an individual asset a is* $q_{at}^D = q_t^D + q_{a,t}^{D,\perp}$, *where* q_t^D *is the demand*

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

change for the aggregate stock market seen in Proposition 4, and $q_{at}^{D,\perp}$ is the asset-a specific demand change, given by

$$q_{at}^{D,\perp} = -\zeta^{\perp} p_{at}^{\perp} + \nu_{at}^{\perp} + \delta\kappa^{\perp} d_{at}^{e,\perp} + \kappa^{\perp}\mathbb{E}_t\left[\Delta p_{a,t+1}^{\perp}\right] \tag{37}$$

where $\zeta^{\perp}$ is the micro-elasticity of demand for individual stocks:

$$\zeta^{\perp} = 1 - \theta^{\perp} + \kappa^{\perp}\delta. \tag{38}$$

Proof. Equation (36) implies

$$q_{at}^{D,\perp} = -\left(1-\theta^{\perp}\right) p_{at}^{\perp} + \kappa^{\perp}\hat{\pi}_{at}^{\perp} + \nu_{at}^{\perp}. \tag{39}$$

Likewise, the analogue of (17) is $\hat{\pi}_{at}^{\perp} = \mathbb{E}_t\left[\Delta p_{a,t+1}^{\perp}\right] + \delta\left(d_{at}^{e,\perp} - p_a^{\perp}\right)$, with $d_{at}^{e,\perp} := \mathbb{E}_t\left[d_{a,t+1}^{\perp}\right]$. Combining the two gives the announced expression. □

This is exactly the same equation as the one for the aggregate stock market, but now in terms of stock-specific deviations. Hence, the economics of the aggregate stock market works for the individual stocks, but in "perpendicular space", i.e. replacing ζ, p_t, q_t^D by $\zeta^{\perp}$, $p_{at}^{\perp}$, $q_{at}^{D,\perp}$, and so on. We next spell this out and draw consequences.

3.6.2 Micro-elasticity of demand versus macro-elasticity of demand

Suppose that there is a "stock specific flow", whereby someone buys $\Delta F_a^{\perp}$ worth of stock a, while selling $\Delta F_a^{\perp}$ of the aggregate stock market, so that the total change in the demand for aggregate stocks is 0. The asset-a specific fractional inflow is $f_a^{\perp} = \frac{\Delta F_a^{\perp}}{P_a Q_a}$, where $P_a Q_a$ is the (pre-flow) market value of stock a. As net demand is 0, we must have $q_{at}^{D,\perp} + f_a^{\perp} = 0$. So, the impact of a flow is:

$$p_a^{\perp} = \frac{f_a^{\perp}}{\zeta^{\perp}}, \tag{40}$$

where $\zeta^{\perp}$ is the price micro-elasticity of demand (38). We see that the price impact is $\frac{1}{\zeta^{\perp}}$, not $\frac{1}{\zeta}$.

Calibration Most papers have estimated the micro-elasticity of demand, $\zeta^{\perp}$ (Shleifer (1986), Wurgler and Zhuravskaya (2002), Duffie (2010), Chang et al. (2014), Koijen and Yogo (2019)), while the present paper is about the macro-elasticity of demand, ζ. Indeed, the literature finds $\zeta^{\perp} \simeq 1$, with estimates in the 0.5 to 10 range. It makes sense that the macro-elasticity should be much smaller than the micro-elasticity, $\zeta \ll \zeta^{\perp}$. One way to rationalize this is to set $\theta^{\perp} \simeq 0.2$ for the inertia or concern for tracking error term, $\delta = 4\%$, and $\kappa^{\perp} = 5$.

Micro versus macro price impact Taking a micro elasticity $\zeta^{\perp} = 1$ and a macro elasticity $\zeta = 0.2$, this means that if a consumer buys \$1 of a small stock (selling \$1 worth of bonds), the market value of that stock goes up by \$1, and the market value of the aggregate stock market goes up by \$5.

To see this analytically, consider a flow $f_a = \frac{\Delta F_a}{P_a Q_a}$ into just one asset a, which accounts for a fraction ω_a of the total equity capitalization. We do that in the two-period model, so we drop t (this is equivalent to doing that for the infinite-horizon model, but assuming permanent inflows).

The corresponding aggregate flow is $f = \omega_a f_a$, so that the impact on the aggregate market is $p = \frac{f}{\zeta}$, or

$$p = \frac{\omega_a f_a}{\zeta}.$$

The stock-specific flow to asset a is $f_a^{\perp} = f_a - f = (1 - \omega_a) f_a$. Hence, the stock-specific impact is: $p_a^{\perp} = \frac{f_a^{\perp}}{\zeta^{\perp}} = \frac{1-\omega_a}{\zeta^{\perp}} f_a$. Hence, the total impact is $p_a = p + p_a^{\perp}$, or

$$p_a = \frac{f_a}{\zeta^{\perp}} + \left(\frac{1}{\zeta} - \frac{1}{\zeta^{\perp}} \right) \omega_a f_a. \tag{41}$$

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

$$p_b = \left(\frac{1}{\zeta} - \frac{1}{\zeta^{\perp}} \right) \omega_a f_a, \qquad b \neq a. \tag{42}$$

As $\zeta < \zeta^{\perp}$, the cross-impact is positive.[33]

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:

$$p_{a,t}^{\perp} = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho^{\perp}}{(1+\rho^{\perp})^{\tau-t+1}} \left(\frac{f_{a\tau}^{\perp} + \nu_{a\tau}^{\perp}}{\zeta^{\perp}} + M^{D,\perp} d_{a\tau}^{e\perp} \right). \tag{43}$$

Conclusion: Aggregate versus cross-section We conclude that the aggregate model transposes well to the cross-section. While most prior work has been on the estimation of the cross-sectional elasticity $\zeta^{\perp}$, the main object of interest in this study is the aggregate elasticity ζ.

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

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

"For every buyer there is a seller" 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"). The above two-period model clarifies that this reasoning is incorrect. In Proposition 2, f is the pressure to buy stocks (if it is positive), and the demand $q^D = -\zeta p + f$ has a component $-\zeta p$ expressing that "sellers" will sell the shares to "buyers" represented by f. So there are both buyers and sellers (or

[33] The logic directly extends to a setup where stocks are separated into disjoined groups, like industries or characteristics-based portfolios. We have for the aggregate that $p = \frac{f}{\zeta}$. Now, for each group G, we have: $p_G - p = \frac{f_G - f}{\zeta^{\perp}}$. And for a stock a in a group G, we have: $p_a - p_G = \frac{f_a - f_G}{\zeta^{\perp}}$. The model can be refined with a group-specific elasticity $\zeta_G^{\perp}$ lower than the stock-specific elasticity $\zeta_a^{\perp}$.

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).

"Fast and smart investors (perhaps hedge funds) will provide elasticity to the market" This is not true: in part because they are small (see Section 2.2), these investors cannot provide 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, smart-money investors often face risk constraints that limit their ability to aggressively step in during aggregate downturns.

"Permanent price impact must reflect information" In Proposition 5, a one-time, non mean-reverting inflow permanently changes prices (as in $\Delta p_0 = \frac{\Delta f_0}{\zeta}$), even if it contains no information whatsoever. Hence, the typical empirical strategy to look for reversals as signs of flows 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.

"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). In our model, these trades within the universe of equities do no count toward the aggregate flow f, although they represent a large cross-sectional flow $f_a^{\perp}$. As we mentioned above, the sum of the absolute values of sector-level flows, including net issuances, relative to the size of the market, averages only 1.9% per year.

"The market often looks impressively efficient in the short turn, 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-elastic (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 $(1+\rho)^{-T} = 99.8\%$ of the eventual present value of the future dividends or inflows (taking the ρ calibrated in Section 3.4 and $T = 1/52$ years). 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 smooth well very persistent shocks.

Furthermore, even though prices respond promptly around major events, it is generally hard to assess whether the market moved by just the right amount, or instead under- or over-reacted. In addition to a large literature demonstrating drifts in prices before and after macro events (such as Federal Open Market Committee meetings), our model implies that persistent flows around such events can lead to persistent deviations in prices, and typical event study graphs that do not display much of a drift in prices following the event would be uninformative about macro efficiency.

"Prices are the discounted present value of expected future dividends" In Proposition 5, prices are the discounted value of both future dividends, demand shocks, *and future flows*.[34] This notion is analogous to the Campbell and Shiller (1988b) decomposition. We refine this decomposition by tracing back changes in future discount rates to change in future flows and demand shocks by investor – which are arguably more directly tangible. By understanding which investors determine discount rate variation, we can potentially test rational or behavioral models of time-varying discount rates more directly.

"Share buybacks do not affect equity returns, as proved by the Modigliani-Miller theorem" In the traditional frictionless model, the price impact of a share buyback should be zero. However, in our model, if firms in the aggregate buy back $1 worth of equity, that increases aggregate valuations — as should be clear, and is detailed below in Sections 5.2 and 12.4. Hence, share buybacks are potentially a large source of fluctuations in the market. In our model, a combination of fund mandates and consumers' bounded rationality leads to a violation of the Modigliani-Miller neutrality.

More broadly, corporate actions such as share issuances, transactions by insiders, et cetera, may have a large impact on prices beyond any informational channel. Most extant empirical evidence focuses on announcements at the firm level, while we emphasize their impact in aggregate. By focusing on well-identified firm-level responses, one identifies $\zeta^{\perp}$, not ζ. It will be interesting to explore in detail how important corporate decisions are for fluctuations in the aggregate stock market.

4 Estimating the Aggregate Market Elasticity

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

In Section 4.2, we provide the basic intuition behind the GIV estimator. We demonstrate the performance of the GIV estimator in our specific setting in Section 4.3 using simulations. We report the estimates in Section 4.4 using sector-level data from the Flow of Funds and in Section 4.5 using investor-level data 13F filings.

In Section 4.6, we trace the volatility of aggregate stock market returns back to different categories of investors, and we highlight in particular the role of sector-specific shocks. We explore the connection between capital flows and asset prices in Section 4.7. We also connect capital flows to macro-economic variables and measures of beliefs to provide an initial analysis of the potential determinants of flows into the equity market.

[34]This effect for stocks has counterpart in bonds (Greenwood and Vayanos (2014)) and exchange rates (Gabaix and Maggiori (2015)).

4.1 Data sources

We combine data from various sources that we briefly outline here; we leave a detailed description for Appendix Section 9. The sample period used varies depending on the data source, and we discuss our choices in terms of data construction and sample selection in the appendix as well.

In short, we combine sector-level holdings data from the U.S. Flow of Funds (FoF), equity data from CRSP, data on firm fundamentals from Compustat, 13F investor-level holdings data from FactSet,[35] detailed portfolio data for mutual funds and exchange-traded funds (ETFs) from Morningstar, detailed portfolio data for state and local pension funds from the Census ASPP and QSPP, macro data from the US Bureau of Economic Analysis, and survey data on return expectations from Gallup.

4.2 Intuition behind the GIV estimator

We estimate the macro multiplier, M, using the GIV approach developed in Gabaix and Koijen (2020). We provide a brief introduction to the methodology in Section 8 and explain the logic in this section. We leave the technical details of the specific algorithms that we use to Appendix Section 8.2.1.

Recall that our model implies, for a given sector j,

$$\Delta q_{jt} = \alpha_j - \zeta \Delta p_t + f_{jt}^{\nu}, \tag{44}$$

where we assume for now that all sectors have the same demand elasticity ζ,[36] and $f_{jt}^{\nu} := f_{jt} + \nu_{jt}$ captures the combined impact of capital flows and demand shocks on the change in demand. Here, Δq_{jt} denotes the relative change in equity holdings, Δp_t the return on the aggregate stock market including dividends, and f_{jt} the per-period inflow into sector j. We additionally assume that flows and demand shocks are driven by real GDP growth, Δy_t, several common factors, η_t, as well as idiosyncratic shocks u_{jt}:

$$f_{jt}^{\nu} = \beta_j \Delta y_t + \lambda_j' \eta_t + u_{jt},$$

where the cross-sectional covariance matrix $V(u_t) = V_u$ is diagonal. Importantly, we assume that the idiosyncratic shocks u_{jt} and the aggregate shocks η_t are uncorrelated, $\mathbb{E}[\eta_t u_t'] = 0$. Similarly, for the corporate sector, we assume that net repurchases are given by

$$\Delta q_t^C = \alpha_C - \zeta_C \Delta p_t + \beta_C \Delta y_t + \lambda_C' \eta_t + u_{Ct},$$

and we drop superscripts for now. We index the corporate sector with $j = 0$. We define the size weight of sector j as the fraction of the total equity market it holds; and we let $S_0 = 1$. The size weights, as a result, add up to two, $\sum_{j=0}^{N} S_j = 2$, and to one for the investor sectors, $\sum_{j=1}^{N} S_j = 1$. We define $x_S = \sum_{j=1}^{N} S_j x_j$ for a given quantity x_j.

This model implies (omitting constants):

$$\Delta p_t = M \left(u_{St} + u_{Ct} + (\beta_S + \beta_C) \Delta y_t + (\lambda_S + \lambda_C)' \eta_t \right), \qquad M = \frac{1}{\zeta + \zeta_C}. \tag{45}$$

Our objective in this section is to estimate M, ζ, and ζ_C.

[35] The same data have also been used in Koijen et al. (2019).

[36] We relax this assumption in Section 4.5.

The idea of the GIV estimator is to use sector-specific shocks u_{jt} to the flows as primitive shocks to the system, and see their impact on the price: we estimate the multiplier M as the response of the price to a size-weighted sum u_{St} of idiosyncratic shocks (equation (45)).

Appendix Section 8.2.1 details how we can extract sector-specific shocks, but suppose for now that u_{jt} and u_{St} are available to us. It then follows from (45) that we can simply regress Δp_t on u_{St} and the slope coefficient is equal to $M = \frac{1}{\zeta + \zeta_C}$. 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.[37] Appendix Section 8.2.1 also clarifies how we can directly estimate the elasticities, ζ and ζ_C.

In using a GIV estimator to estimate the multiplier and elasticities, two general threats to identification are that: (i) we do not properly control for common factors; and (ii) the loadings of the omitted factor are correlated with size, such that $\lambda_S - \lambda_E \neq 0$, with λ_E the equal-weighted average. To mitigate the risk of omitted factors, we extract additional factors and explore the stability of the estimates as we add extra factors.

4.3 Performance of the GIV estimator: Simulations

We refer to Gabaix and Koijen (2020) for a more extended analysis of the GIV estimator. Here, we illustrate its performance in an environment that closely mimics our empirical setting.

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 – the micro elasticity) to $\zeta = 0.1$ (a multiplier of 10). For each value of ζ, we construct $f_t^{\nu} = \Delta q_t + \zeta \Delta p_t$. We then assume that the data follow a factor model and estimate $f_t^{\nu} = \lambda \eta_t + u_t$ using principal components analysis. This provides us with an estimate of λ and V_u. These are the estimates one would obtain given the data we observed historically and if the true elasticity were equal to the assumed value. It tells us the volatility of aggregate shocks (μ_λ), the average volatility of idiosyncratic shocks ($\mu_{\ln \sigma}$), and the dispersion in these parameters across investors (σ_λ and $\sigma_{\ln \sigma}$).

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

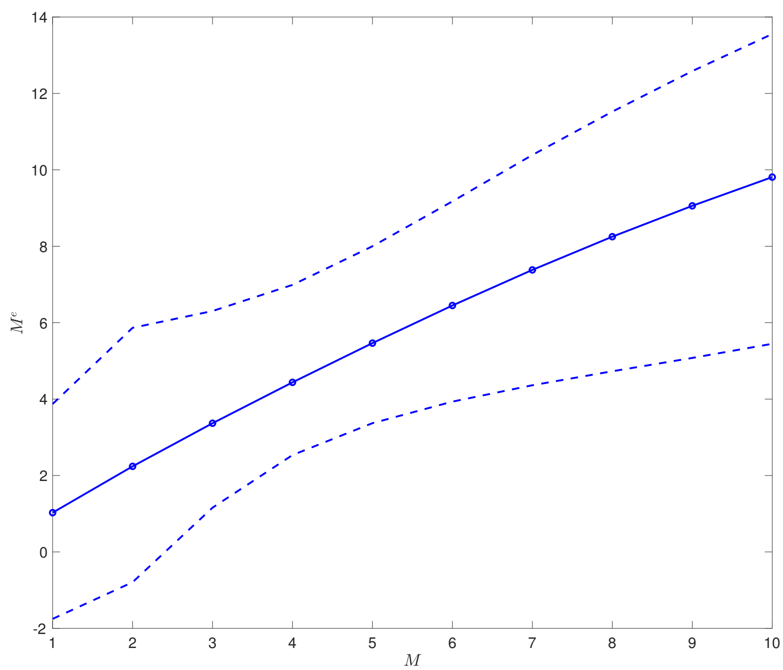
We consider 50,000 replications for each value of $\zeta = 0.1, 0.2, \ldots, 1$ and report the average estimate alongside the 2.5% and 97.5% percentiles across all replications in Figure 3. We report the multiplier M corresponding to the true data-generating process on the horizontal axis and the distribution 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.

4.4 Elasticity estimates using sector-level data

Benchmark estimates We first report the GIV estimates of the macro elasticity using data from 1993.Q1 to 2018.Q4 using the Flow of Funds (FoF). The results are presented in Table 1. The first

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

Figure 3: 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.



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.[38] 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 (90). 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 elasticity of the corporate sector. The short-run elasticity is low and equals 0.01 for both one and two principal components.[39] 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.[40]

In the final column, we report the same regression as in the 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.

Longer-horizon estimates In Figure 4, we explore how demand and flow shocks propagate across time. To this end, we extend the earlier analysis by estimating

$$p_{t+h} - p_{t-1} = a_h + M_h Z_t + c_h \eta_t^e + \lambda_h \Delta y_t + \epsilon_{t+h}, \tag{46}$$

for $h = 0, 1, \ldots, 4$ quarters. Recall that η_t^e is the principal component, extracted in the fourth step of the GIV procedure as outlined in Section 8.2.1. The figure reports M_h at a certain horizon. We also consider a regression where we replace the left-hand side by $p_{t-1} - p_{t-2}$, to which we refer as $h = -1$. To construct the confidence intervals, we account for the autocorrelation in the residuals due to overlapping data.

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

Robustness We explore the robustness of our estimates in various dimension. In the interest of space, we report the tables in Appendix Section 10.4 and summarize the results here.

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

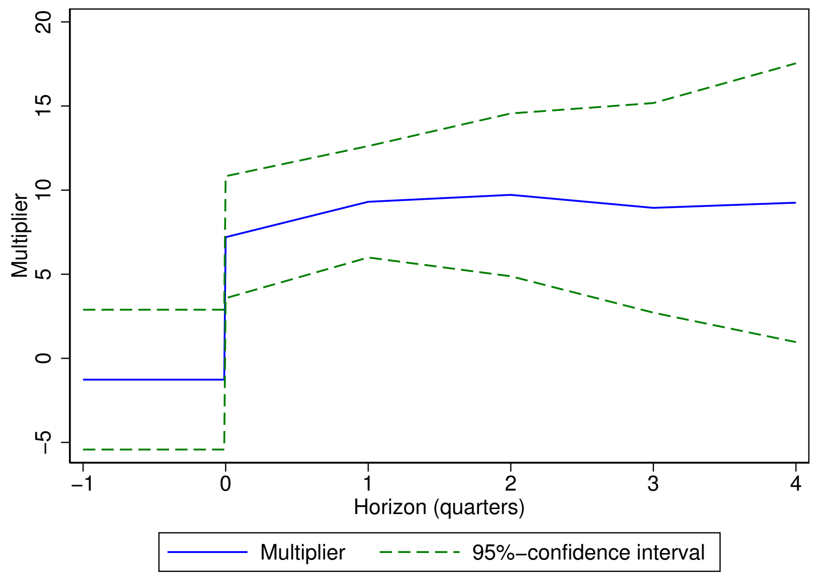
[39] This small contemporaneous elasticity of the supply of shares by the corporate sector, estimated here by GIV, is consistent with the OLS findings of Ma (2019). She finds (Table VII) that $\frac{\text{Gross equity issuance}}{\text{Assets}} = 0.01\hat{\pi}$ (plus other terms) at the quarterly frequency. 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 17) $\zeta_C = \delta \times 0.06 = 0.0024$.

[40] These estimates imply that the short-run elasticity of the corporate sector is low. However, they 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 1: Estimates of the macro elasticity. The first two columns report estimates of M with one and two principal components, η_i, respectively. The next two columns report the elasticity estimates, ζ. 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 to estimate the impact of sector-specific shocks on prices. The sample is from 1993.Q1 to 2018.Q4. The final line reports the R-squared values for the OLS regressions.

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

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 price from $t-1$ to $t+h$ by Mf_t times 1%. We use the GIV for instrumentation, see (46). 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.



In Table 4, we consider the impact of not merging mutual funds and ETFs, not adjusting for principal components, not including time trends in the models for Δq_{jt}, controlling for lagged Δq_{jt}, and using the sample from 2000.Q1 to 2018.Q4. The main takeaway is that the multiplier estimates are stable across the various specifications. The estimates of the multiplier vary between 6.6 and 8.0.

In Table 5, we replicate Table 1, but now adding the lagged value of Z_t. Across various specifications, the GIV has a small (insignificant) positive autocorrelation of approximately 10-15%. As the autocorrelation is so minimal, it does not affect our estimates.

Lastly, in Table 6, 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. A reasonable concern is therefore that we miss an aggregate factor. We present the results up to five principal components for completeness. We find a multiplier of 5.3 with two principal components, 3.9 with three principal components, 4.2 with four principal components, and 3.5 with five principal components. As our main concern is the small cross-section, we turn to investor-level data in the next section as an alternative approach to estimating the multiplier.

4.5 Elasticity estimates using investor-level data

We repeat the same exercise but now using more disaggregated, investor-level 13F data.[41] We use 13F data from FactSet that covers the period from 1999.Q2 to 2017.Q2. The 13F data has two potential advantages over the FoF data. First, we have a large cross-section of investors, which allows us to accurately extract latent factors. Second, we can construct investor-level characteristics, such as their size, turnover, and characteristics of their investment strategy (e.g., with respect to size, value, and market beta). This allows us to estimate the model with heterogeneous demand elasticities across investors and uncover common factors associated with those characteristics (e.g., common flows from value- to growth-oriented, technology funds in the late nineties). The disadvantage, however, is that we do not have fixed income holdings and we cannot make the connection to capital flows. In addition, the sample is shorter.[42] We therefore use the 13F data as another opportunity to estimate the elasticity using a different level of aggregation.

In this section, we start from our model as before:

$$\Delta q_{jt} = -\zeta_{j,t-1}\Delta p_t + \beta_j \Delta y_t + \lambda'_{j,t-1}\eta_t + u_{jt},$$

where we allow for heterogeneity in demand elasticities, $\zeta_{j,t-1}$. We assume a parametric specification for elasticities and a semi-parametric specification for factor loadings:

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

where $x_{j,t-1}$ is a vector of investor characteristics of which the first element is equal to 1, and $\dot{\zeta}$, $\dot{\lambda}$, and $\ddot{\lambda}_j$ are to be estimated. As investor characteristics, we use log turnover, active share, and log size, as well as portfolio characteristics, namely the value-weighted market capitalization, log book-to-market ratio, and market beta, to estimate the factors.[43] The macro elasticities can also be heterogeneous across investors and depend on the same, or a subset of the, characteristics. In addition, we allow for non-parametric factors via $\ddot{\lambda}_j$.

Using this structure, we implement the GIV procedure from Gabaix and Koijen (2020), that we detail in Section 8.2.2. We also provide further details on measuring flows, sample selection criteria, and how we construct the characteristics in Appendix Section 9.3. The results are presented in Panel A of Table 2. In the first column, we only use parametric factors, that is, $\ddot{\lambda}_j = 0$. We estimate a multiplier of $M = 5.5$ with a standard error of 3.2. In columns 2 to 6, we add one to five principal components, respectively. This is in addition to six parametric factors associated with the characteristics and a common factor on which all investors load equally. The multiplier varies between $M = 4.6$ and $M = 7.2$.

To conclude, we also implement an over-identifying restrictions test as suggested by Gabaix and Koijen (2020). In particular, after estimating the idiosyncratic shocks, u_{jt}, we construct two GIV instruments. Each period, we rank all funds by their lagged size and we put the funds with an odd rank in one group and those with an even rank in the other group. We then construct a separate

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

[42] Alternatively, we can use data from Thomson Reuters that has a history going back to 1980. However, Thomson Reuters has several measurement issues that can lead to some large flows, which is undesirable for the GIV estimation.

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

Table 2: Multiplier estimates using 13F data. The table reports the multiplier estimates using 13F data. The benchmark results are presented in Table 2. In the first column, we only use parametric factors. The characteristics used are log asset under management, active share, log turnover as well as the value-weighted size of the portfolio, log market-to-book ratio, and market beta. In the next five columns we add one to five principal components, respectively. All regressions include a constant that we omit for brevity. In Panel B, we test for over-identifying restrictions by constructing two GIV instruments. After extracting the idiosyncratic shocks, we sort funds each period by their lagged size and split them into an odd and even group based on their ranks. The sum of the instruments is the instrument used in Panel A, and we regress returns on z, the difference in instruments, and GDP growth. Under the null of the model, the coefficient on z_t should be the same as in Panel A and the coefficient on $z_t^{Odd} - z_t^{Even}$ should be zero. In Panel C, we regress the instrument based on the odd observations on the instrument based on the even observations. Under the null of the model, the regression coefficient should be zero. The sample is from 1999.Q2 to 2017.Q2.

Panel A: Benchmark results

Number of principal components	0	1	2	3	4	5
z	5.53	5.17	5.71	4.57	6.91	7.21
	(3.18)	(3.61)	(2.61)	(3.20)	(3.06)	(2.83)
GDP growth	6.34	6.29	6.39	6.35	6.37	6.41
	(0.89)	(0.87)	(1.10)	(1.08)	(1.21)	(1.35)
R^2	0.230	0.225	0.232	0.216	0.242	0.245
Observations	72	72	72	72	72	72

Panel B: Over-identifying restrictions test

Number of principal components	0	1	2	3	4	5
z	5.04	4.89	5.33	4.08	6.63	6.74
	(3.36)	(3.48)	(2.79)	(3.08)	(3.42)	(3.33)
$z^{Odd} - z^{Even}$	2.53	1.77	2.06	1.91	1.06	1.71
	(3.40)	(3.83)	(3.88)	(3.41)	(4.04)	(4.06)
GDP growth	6.33	6.29	6.40	6.32	6.37	6.42
	(0.94)	(1.08)	(1.11)	(1.52)	(0.81)	(0.70)
R^2	0.239	0.229	0.238	0.220	0.243	0.248
Observations	72	72	72	72	72	72

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

Number of principal components	0	1	2	3	4	5
z^{Even}	-0.11	-0.11	-0.08	-0.12	-0.11	-0.15
	(0.11)	(0.09)	(0.08)	(0.09)	(0.12)	(0.13)
R^2	0.009	0.009	0.005	0.009	0.008	0.014
Observations	72	72	72	72	72	72

instrument for the odd group, z_t^{Odd}, and for the even group, z_t^{Even}. By design, $z_t = z_t^{Odd} + z_t^{Even}$. We also construct their difference, $z_t^{Odd} - z_t^{Even}$.

We then run the final regression

$$\Delta p_t = \alpha + M z_t + M^{Odd-Even}\left(z_t^{Odd} - z_t^{Even}\right) + \delta \Delta y_t + \epsilon_t,$$

and we test whether $H_0 : M^{Odd-Even} = 0$. The results are presented in Panel B of Table 2. We find that the estimates of M are very close to Panel A and the coefficient on the difference is economically quite small and statistically insignificant. Of course, given the fairly short sample, the standard errors are large.

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

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

4.6 Decomposing fluctuations in the aggregate stock market

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

$$s_j^p = \frac{Cov\left(\Delta p_t, \frac{1}{\zeta} S_{j,t-1} f_{jt}^{\nu}\right)}{Var\left(\Delta p_t\right)}, \tag{47}$$

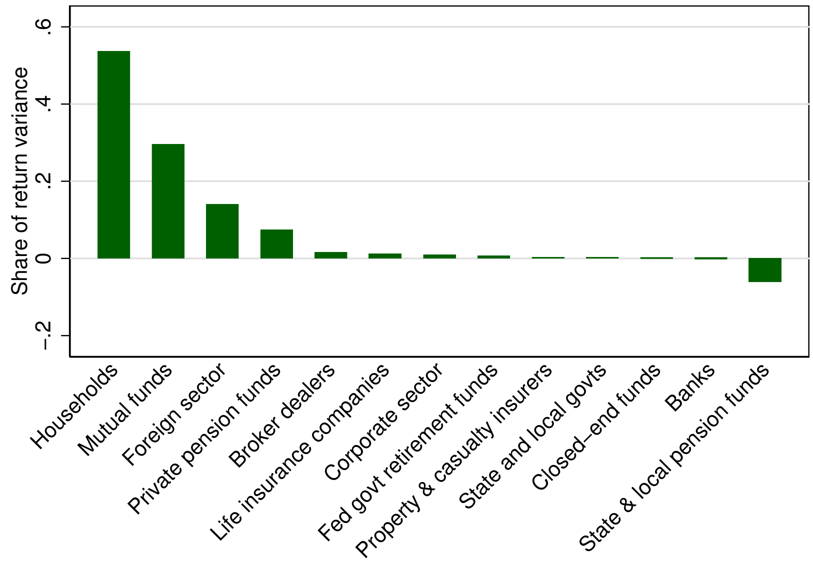
which sums to 100% across investors as we include the corporate sector, where $\frac{1}{\zeta} S_{0,t-1} f_{0t}^{\nu} = \frac{1}{\zeta} f_{Ct}^{\nu}$. We report the variance shares in Figure 5 for $\zeta = 0.13$, based on the third column of Table 1. We find that the household sector accounts for 49% of the variation in stock returns, followed by mutual funds and ETFs (28%), the foreign sector (14%), and private pension funds (8%). The somewhat negative contribution of state and local pension funds may reflect once again their somewhat higher demand elasticity. The share of variance is different from the equity share presented in Figure 1.

In addition, we assumed $f_{St}^{\nu} = \beta_S \Delta y_t + \lambda_S' \eta_t + u_{St}$ and we can compute the importance of idiosyncratic sector-level shocks:

$$\frac{Var\left(u_{St}\right)}{Var\left(f_{St}^{\nu}\right)}. \tag{48}$$

To this end, we run a panel regression of Δq_{jt} on time fixed effects, sector fixed effects, and GDP growth (with heterogeneous slopes). We collect the residuals and extract another principal component, as in the third step of the GIV algorithm outlined in Section 8.2.1. We use the residuals to compute u_{St} and compare their variances. In Table 7 in the Online Appendix, we also report the volatility of sector-specific demand shocks by sector. Using these estimates, the contribution of

Figure 5: Contribution to variance. The figure reports the contribution to the variance of the US equity market of different sectors. The variance shares are computed using equation (47) with $\zeta = 0.2$. This may be compared to the "size shares" in Figure 1.



idiosyncratic sector-specific shocks as in (48) is 20%.[44] In the next section, we further decompose f_{jt}^{ν} into net flows f_{jt} and demand shocks ν_{jt}.

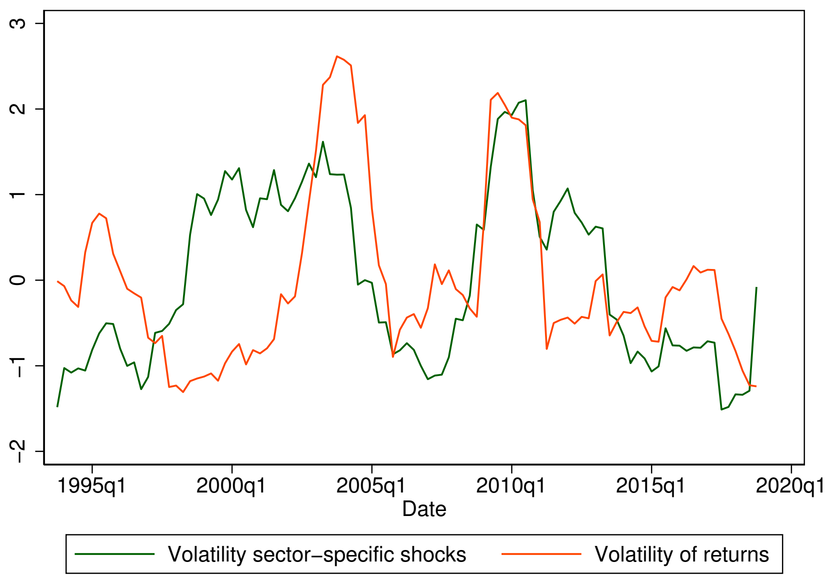
An important feature of stock return data is that the volatility changes over time. Sector-specific shocks are not only relevant for the unconditional volatility of the aggregate stock market, but also for its time variation. We compute the two-year rolling standard deviation of sector-specific idiosyncratic shocks, u_{jt}. We then compute $\sigma_{u_{St}} = \left(\sum_j S_{j,t-1}^2 \sigma_{u_i,t}^2\right)^{\frac{1}{2}}$ and we also compute the rolling standard deviation for stock returns. Figure 6 plots both series (both series are standardized), which track each other closely. Their time series correlation is 41%. This implies that sector-specific shocks are important not only to understand the unconditional volatility of the market, but also its time variation.

4.7 Aggregate capital flows, prices, macro-economic variables, and beliefs

Measuring flows So far, we have studied f_{jt}^{ν}, which is the sum of capital flows, f_{jt}, and demand shocks, ν_{jt}. In this section, we measure capital flows directly and study their properties. We first discuss the measurement of flows in a way that is consistent with our theory. This provides us with a measure of aggregate capital flows into the equity market. As this measure is new to the literature, we show its connection to prices, macro-economic variables, and beliefs. We emphasize that this just illustrates a correlation and should not be interpreted causally — unlike the previous part of this Section 4, which were concerned with causality and identification. The results in this

[44] If we take out a sector-specific time trend, as in (87), and interpret the trend as a common factor, the variance share declines to 16%.

Figure 6: Volatility of flows and volatility of returns.. We compute the volatility of idiosyncratic demand shocks by the various sectors (u_{St}) using a two-year rolling average and we do the same for the return on the aggregate stock market. The sample is from 1994.Q1 to 2018.Q4. This is consistent with the model, where a high volatility of flows implies a high volatility of returns.



section provide an initial analysis of the potential determinants of flows into the aggregate stock market.

Our measurement is guided by our theory. As discussed in Section 3, to measure flows into the market, we cannot look at equity purchases or standard measures of flows. After all, $\sum_{j=0}^{N} S_j \Delta q_{jt} = 0$, where sector $j = 0$ corresponds to the corporate sector, by market clearing. To measure flows correctly, we need to look at the *total flow* of an investor into stock and bond markets. In this section, we measure the bond flow as the combined flow into Treasury and corporate bond markets.

Ideally we would know the equity and fixed income holdings of investors, but such data are not available for all institutions.[45] We therefore rely on the FoF for these calculations, and we refer to Appendix Section 9 for details on the data construction. However, the FoF aggregates data across many institutions, and the reported flows can be mismeasured. To see this, consider the case in which some households only invest in bonds and other households only invest in equities. If we view this as a combined household, a 1% combined inflow into financial markets does not necessarily lead to a 1% increase in equity holdings as the flow may be a flow to bond funds only.[46] With disaggregated data, such problems can be solved, but such data are unfortunately unavailable.[47]

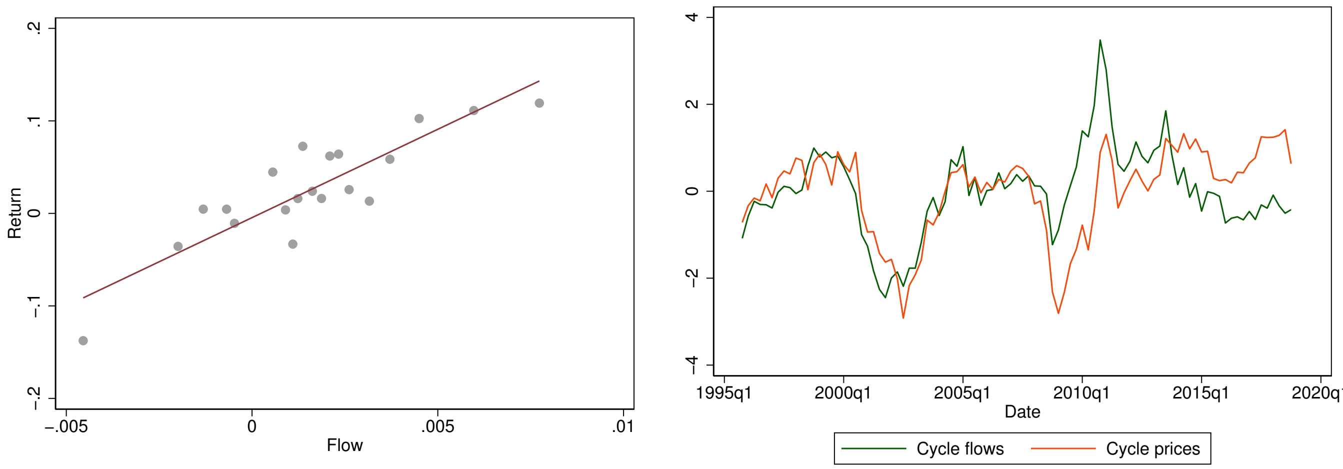
We propose a simple diagnostic to assess whether flows are measured accurately. In particular,

[45] While several institutions, like mutual funds and insurance companies need to report their fixed income holdings, the same is not true for many other investors. That is, there is no 13F counterpart for fixed income securities, unfortunately.

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

[47] In the Euro area, supervisory data, namely the Security Holdings Statistics (SHS), are available in recent years that would allow for such an analysis.

Figure 7: Capital flows into the stock market and price changes. We plot the aggregate flow into the stock market, $f_{St} = \sum_{j=0}^{N} S_j f_{jt}$, versus the return on the aggregate stock market in the left panel used a binned scatter plot. In the right panel, we construct a cumulative (log) return index and compute cumulative flows. We extract the cyclical component using the methodology developed in Hamilton (2018). We standardize both measures over the full sample to be able to plot them in the same figure. The sample for both figures is from 1993.Q1 to 2018.Q4.



in our model, the elasticity of demand to flows equals one, see (7). We therefore estimate

$$\Delta q_{jt} = \alpha + \beta_j f_{jt} + \gamma_j \Delta p_t + \delta_j' x_{jt} + \epsilon_{jt}, \tag{49}$$

where x_{jt} includes $f_{j,t-1}$ and Δy_t. We only include the flows for which we cannot reject the null hypothesis $H_0 : \beta_j = 1$. We report the estimates of β_j in Table 8 in Appendix Section 10. When we cannot reject $H_0 : \beta_j = 1$, we use the total flow and for the other sectors we replace it with the equity flow.[48]

Flows into equities and equity returns Next, we relate our measure of capital flows into the stock market to returns. In the left panel of Figure 7, 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 these are merely correlations and likely biased upward if flows respond positively to prices within a quarter.[49]

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

[48]Alternatively, we can use a cutoff of $|\beta_j - 1| < 0.3$, for instance. The cutoff of 0.3 is obviously arbitrary and changes in this cutoff do not matter much (e.g., 0.2 or 0.4) until the household sector is included (the household sector has a slope coefficient of $\beta_j = 0.45$, see Table 8). This sector is large and flows do not explain equity purchases well.

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

Table 3: 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 normalized them to have unit standard deviation. In the next four columns, 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 innov	Flow innov	Flow innov	Return	Return	Return	Return
Gallup innov	0.48		0.46		0.61		0.33
	(0.10)		(0.11)		(0.09)		(0.09)
GDP growth innov		0.21	0.06			0.41	0.21
		(0.11)	(0.11)			(0.10)	(0.08)
Flow innov				0.65			0.45
				(0.09)			(0.09)
Constant	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00
	(0.10)	(0.11)	(0.10)	(0.09)	(0.09)	(0.10)	(0.07)
Observations	79	79	79	79	79	79	79
R^2	0.233	0.046	0.237	0.426	0.376	0.171	0.582

then extract the cyclical component using the methodology developed in Hamilton (2018). We standardize both measures over the full sample to be able to plot them in the same figure. These are shown in the right panel of Figure 7. Consistent with the high-frequency co-movement that we uncover in the left panel of Figure 7, we find that prices and flows co-move at a business cycle frequency.

We re-emphasize once again that these are merely correlations. It may be the case that they reflect positive feedback trading by investors. We do note, however, that positive feedback trading makes the overall market more inelastic as the demand curve implied by flows is upward sloping.

Relating flows to shocks to GDP and to 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 Section 9. 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. We standardize each of the innovation series to simplify the interpretations of the regressions.

The results are presented in Table 3. In the first three columns, we relate capital flows to survey expectations and economic growth. We find that flows and survey expectations are strongly correlated, confirming Greenwood and Shleifer (2014). A one standard deviation increase in survey expectations of returns is associated with a 0.48 standard deviation increase in capital flows. This finding may resolve a recent challenge posed to the beliefs literature by Giglio et al. (2019). In particular, they find that although survey expectations of returns are volatile, the passthrough 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 passthrough is low, but small flows into inelastic markets lead to large price effects.

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

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

5 Finance with Inelastic Markets: Government and Corporate Policies, Pricing Kernel, Endogenizing Flows, Other Asset Classes

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

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

This paper suggests that the government might prop up asset values, perhaps in times of crisis, or to help firms invest by raising equity at a high price.

Indeed, suppose that the government buys f^G percent of the market, and keeps it forever. Then, the market's valuation increases by $p = \frac{f^G}{\zeta} \simeq 5f^G$.[50] So, if the government buys 1% of the market (which may represent roughly 1% of GDP), the market goes up by 5%. If the government buys it for just T periods, the impact is $p = \left(1 - \frac{1}{(1+\rho)^T}\right) \frac{f^G}{\zeta}$.[51] With the above calibration, this can be a moderate dampening if T is large enough.

This is what a number of central banks have done. In August 1998, the Hong Kong government, under speculative attack, bought 6% of the Hong Kong stock market: this resulted in a 24% abnormal return, which was not reversed in the following eight weeks (Bhanot and Kadapakkam

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

[51] Set $f_t = f^G 1_{0 \leq t < T}$ in (20).

(2006)). This effect is not entirely well-identified, but is consistent with a large price impact multiplier $\frac{1}{\zeta}$, perhaps 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.[52] In inelastic markets, this may have a large price impact. It would be interesting to quantify the macro elasticity of those markets, and to develop a normative theory of these government purchases of equities.[53]

5.2 Corporate finance in inelastic markets

Imagine that the average firm buys back shares in one period, reducing dividends and hence keeping total payouts constant. What happens?

In a frictionless model, this does not affect the firm's value (Modigliani-Miller is at work). In an inelastic model, it should now be clear that buybacks will increase the firm's value. We now detail this.

For clarity and brevity, we focus on the two-period model (the same economics holds in infinite horizon, but the expressions are more complicated; see Section 12.4). 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 $Q_0' = Q_0(1-b)$). The buyback is financed by a fall in the time-0 dividend, so the total dividend payout falls from $\mathcal{D}_0$ to $\mathcal{D}_0' = \mathcal{D}_0 - P_0 Q_0 b$, where P_0 is the ex-dividend price, and $P_0 Q_0 b$ is used to finance the share buyback.

We need to take a stand on the households' reaction to those buybacks. Call μ^D (respectively μ^G) the fraction of the change in dividends (respectively, of the change in capital gain) that is "absorbed" by the households – that is, consumed or reinvested in the pure bond fund. If the extra dividend (respectively extra capital gain) is X dollars, consumers will "remove from the mixed fund" $\mu^D X$ (respectively $\mu^G X$) dollars. As households' marginal propensity to consume is higher after a \$1 dividend rather than a \$1 capital gains (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 mental accounting. We simply trace their implications for the price impact of share buybacks in the following proposition (which is proved in Section 11).

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

$$v = \frac{\left(\mu^D - \mu^G\right)\theta}{\zeta + \mu^G \theta} b, \tag{50}$$

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: $v > 0$.

[52] Lockett, Hudson. "How the invisible hand of the state works in Chinese stocks." *Financial Times*, February 4, 2020. Accessed June 17, 2020. https://www.ft.com/content/0d41cb6e-4717-11ea-aeb3-955839e06441

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

A very tentative calibration Using the recent estimates of Di Maggio et al. (2018), we set $\mu^G \simeq 0.03$, and $\mu^D \simeq 0.5$.[54] This calibration gives (with θ as in Section 3.4) the value $\upsilon \simeq 3b$ for (50). Hence, under this tentative calibration, a buyback of 1% of the market increases the market capitalization by 1.8%. The generalization to an infinite horizon (Proposition 15) confirms this calibration, as an order of magnitude.

5.3 Pricing kernel consistent with flow-based pricing

Much of asset pricing uses pricing kernels, or stochastic discount factors (SDF). 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 We first define the basic notions. 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:

$$\mathcal{M}^d = \frac{1}{R_f}. \tag{51}$$

In the spirit of maintaining a continuity with the heritage of Lucas (1978), we 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 12.7.

From default pricing kernel to actual pricing kernel

The default pricing kernel $\mathcal{M}^d$ will not price assets correctly, as it does not react to flows. We propose a method of "pricing kernel completion" that will augment the pricing kernel so that it correctly prices all assets. We posit the existence of a very small mass ε (which we will take to be infinitesimal, so that it won't impact prices) of "agile optimizers", who start with zero financial wealth and whose objective function is:

$$\max_Q \mathbb{E}\left[-\mathcal{M}^d e^{-Q'R}\right], \tag{52}$$

where R is the vector of excess returns at time 1.[55] 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 $\mathcal{M}^d$.[56] Hence we have $\mathbb{E}\left[\mathcal{M}^d e^{-Q'R} R\right] = 0$. So, the following $\mathcal{M}$ is a pricing kernel:

$$\mathcal{M} = \mathcal{M}^d e^{-Q'R+\xi}, \tag{53}$$

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

[55] The implicit risk aversion of 1 is just a normalization.

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

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 (53) 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 $\mathcal{M}^{d,R_f} = \frac{1}{R_f}$. Section 12.7 expands to multiple risky assets.

Flow-based SDF for the two-period model Let us revisit the two-period model of Section 3.1. The excess risk premium is $\hat{\pi} = \delta(d - p)$ with p given in (12), so that, with $f = (1-\theta)d + \tilde{f}$, the total risk premium is: $\pi = \bar{\pi} - \delta\frac{\tilde{f}}{\zeta}$. So, the completed pricing kernel is:

$$\mathcal{M} = \exp\left(-r_f - \pi\frac{\varepsilon^d}{\sigma_d^2} + \xi\right), \qquad \pi = \bar{\pi} - \delta\frac{\tilde{f}}{\zeta}, \tag{54}$$

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 (54) just reflects that. If there is a flow f, that modifies the price P according to (12), and the pricing kernel $\mathcal{M}$, in such a way that $P = \bar{P}(1+p) = \mathbb{E}[\mathcal{M}D]$ holds.[57] The pricing kernel is in a sense a symptom rather than a cause in that market.

Flow-based SDF for the infinite-horizon model We next turn to the infinite-horizon model process for flows in (26)-(27), something also delivered by our general equilibrium model of Section 10. The completed SDF is:[58]

$$\mathcal{M}_{t+1} = \exp\left(-r_f - \pi_t\frac{\varepsilon^d_{t+1}}{\sigma_d^2} + \xi_t\right), \qquad \pi_t = \bar{\pi} + b_f^{\pi}\tilde{f}_t \tag{55}$$

where $\xi_t = -\frac{\pi_t^2}{2\sigma_y^2}$ if ε^d_{t+1} is Gaussian. The risk premium is given in (27), and hence depends on flows. Again, flows are central, and the SDF (55) just reflects them after the fact, much as a book-keeping device rather than a primitive source of pricing.

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.

[57] It is a good exercise to verify this directly, with $p = \frac{f}{\zeta} + \mathcal{M}^D d$ as in (12).

[58] 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 risk premium π_t independent of maturity. So, the maximum Sharpe ratio is achieved via a one-period dividend strip.

5.4 Different asset classes

We can easily extend the model to K asset classes, indexed by $A \in \{1, \ldots, K\}$, such as stocks, long-term government bonds, and long-term corporate bonds. This way, we can study cross-market contagion effects, and the impact of those on real investment. We sketch this for the two-period model of Section 3.1.

The mandate leads to the following demand for asset A (at least, for some small deviations from 0 in d and p):

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

For instance, if $\kappa_{AB}^D = 0$ 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 A class changes by a fraction

$$q_A^D = -\sum_B \zeta_{AB} p_B + f_A + \sum_B \kappa_{AB}^D d_B, \tag{56}$$

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

$$q = -\zeta p + f + \kappa^D d, \tag{57}$$

where now q, p, d, f are vectors, and ζ and κ^D are matrices, with dimension K. This generalizes Proposition 2. So, the equilibrium after a change in flows and expected dividends (but still constant asset supply is):

$$p = \zeta^{-1}\left(f + \kappa^D d\right). \tag{58}$$

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.

5.5 Determinants of flows and long-run equilibrium

The present paper mostly points out how impactful flows are. But it is clear that it would be desirable to know more about the determinants of flows at a high frequency. We provided some simple correlations in Section 4.7, but this is clearly a first pass, and those correlations suggest but do not establish causal mechanisms. Establishing the various channels of flows could be a whole line of enquiry, perhaps with micro data such as those used by Calvet et al. (2009) or Giglio et al. (2019). To appreciate the richness of those determinants, let us observe that flow shocks could come from varied sources, such as: (i) changes in beliefs about future flows or fundamentals, as these both affect expected returns, per Proposition 5; (ii) “liquidity needs”, for instance insurance companies selling stocks after a hurricane; (iii) more generally, heterogeneous income or wealth shocks to different groups (including foreign vs 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. (2020); (vii) "road shows" in which firms or governments try to convince potential investors to buy into a prospective equity offering or privatization; (viii) mechanical forced trading via so-called "delta hedging," whereby traders who have sold put options and continuously hedge them need to sell stocks when stock prices fall.

We provide one microfoundation for flows in the macro model of Section 6.1, via the "behavioral disturbance" b_t, which is a stand-in for the above forces. There, the flows are given by (72), which allows us to think about general equilibrium. Here, we examine variants of that formulation. In this section, we wish to highlight a necessary structure of flows in the long run – something that would not be detectable in high quality data over a small sample. We offer pointers to the modeling of flows, and their determinants, in particular in the long run.

A necessary trend and cycle decomposition for flows First, we record that all sensible models of flows should satisfy the following decomposition. By "sensible", we mean models that generate a P/D ratio that is stationary, though potentially stochastic (like an AR(1)).

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

$$f_t = (1-\theta)\,d_t + \hat{f}_t, \tag{59}$$

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.[59]

Proof. Recall (19), $q^D = f_t - (1-\theta)\,p_t + \kappa\hat{\pi}_t$. As $q^D = 0$, (59) holds, with $\hat{f}_t = -\kappa\hat{\pi}_t + (1-\theta)\,(p_t - d_t)$. □

How does the market equilibrate in the long run? One might ask, how does the market discover the trend $(1-\theta)\,d_t$ in (59)? It came out in the model of Section 6.1, 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 12.3 provides details.

Consider a behavioral rule of the type

$$\Delta f_t = \chi\hat{\pi}_t + \varepsilon_t, \tag{60}$$

with $\chi > 0$: this means that people invest more in equities when they are undervalued, which makes flows stabilizing. Then, one can show that this leads to a stationary P/D ratio, as in Lemma 1, and hence the correct representation (59) (see Section 12.3.3).[60]

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. Section 12.3 provides the proof and complements.

[59] The value of $\bar{d}_t$ is only unique "up to a stationary process", but everything works with that caveat in mind.

[60] Rule (60) can have microfoundations of the "behavioral inattention" type:

$$f_t = m f_t^r + (1-m)\,f_{t-1} + \tilde{f}_t, \tag{61}$$

Proposition 9. (Equilibrium when flows respond to the risk premium in a noisy fashion) *In the limit of small time intervals, the specification of flows* (60) *with i.i.d. shocks* ε_t *generates a deviation of the price from trend equal to:*

$$\mathbb{E}_0 p_t = \frac{1}{\zeta + \kappa\phi}\mathbb{E}_0 f_t, \qquad \mathbb{E}_0 f_t = (1-\phi)^t f_0. \tag{62}$$

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 risk 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 (25). This captures that endogenously, flows are "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 $\chi = 0.23$, we replicate a slow mean-reversion of the P/D ratio of $\phi \simeq 9\%$ per year.[61,62]

One could imagine agents with other behavioral rules, or agents optimizing on the parameters χ, m, this way providing additional cross-asset predictions.[63] We leave that to future research.

6 General Equilibrium with Inelastic Markets

So far, we took both the risk-free rate r and the average risk 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.

---

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 $f_t^r = f_t + \frac{\hat{\pi}_t}{\bar{\pi}}$, behavior (61) generates (60) with $\chi = \frac{m}{1-m}\frac{1}{\bar{\pi}} > 0$ and $\varepsilon_t = \frac{\tilde{f}_t}{1-m}$. Formulation (61) extends more easily than (60) to other contexts (Gabaix (2014)).

[61] We compute the dividend yield by summing dividends during the last 12 months relative to the current level of the CRSP value-weighted return index from January 1945 to December 2018. The annual autocorrelation of the log dividend yield during this sample is equal to 0.91 (OLS s.e. 0.048), so we take $1 - \phi = 0.91$.

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

[63] Alternatively, consider a rule like:

$$\Delta f_t = \chi\hat{\pi}_t + \beta(d_t - p_t) + \Delta\tilde{f}_t, \tag{63}$$

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 (63) also leads to the correct form shown in Lemma 1. However, a rule like $f_t = \chi\hat{\pi}_t + \beta(d_t - p_t) + \tilde{f}_t$ would not lead to a stationary P/D ratio: while the right-hand side would be stationary, by Lemma 1 the left-hand side should not be stationary.

6.1 Setup

We consider an infinite horizon economy.[64] For simplicity, we discuss in detail an endowment economy – then, it will be easy to generalize to a production economy. The endowment Y_t follows a proportional growth process, with an i.i.d. lognormal growth rate G_t:

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

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

The representative firm raises capital entirely through equity, and passes the endowment stream as a dividend, $D_t = \frac{Y_t}{Q^{\mathcal{E}}}$. Bonds are in zero net supply.[65] We write the price of equities as

$$P_t = \frac{D_t}{\delta} e^{p_t}, \tag{64}$$

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

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

$$\theta_t = \theta \exp\left(-\kappa^D p_t + \kappa \mathbb{E}_t\left[p_{t+1} - p_t\right]\right), \tag{65}$$

which is the same as before in (3), to the leading order (in terms of deviations from the steady state), with $\kappa^D = \kappa\delta$. The formulation here is slightly more general.[66]

Consumption and investment by households We describe the behavior of the representative household. Section 12.6 supplies more formalism and further details. The dynamic budget constraint on the pure bond fund side is:[67]

$$Q_t^{B,h} + D_t^h = C_t^h + \Delta F_t^h + \frac{Q_{t+1}^{B,h}}{R_{f,t}}. \tag{66}$$

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

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

[64] Section 12.5 develops the model in a two-period economy.

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

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

[67] There is also the usual transversality condition, $\lim_{t\to 0} \beta^t c_{ht}^{-\gamma} Q_t^{B,h} = 0$.

The rational consumer part of the household maximizes lifetime utility, subject to the dynamic budget constraint for bonds (66). She takes the actions of the investor as given.[68] As she is rational, she satisfies the Euler equation for bonds:

$$\mathbb{E}_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}R_{ft}\right]=1, \tag{67}$$

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

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

$$\theta^{M*}=\underset{\theta^M}{\operatorname{argmax}}\,\mathbb{E}\left[V^p\left(\left(1-\theta^M\right)R_{ft}+\theta^M R_{M,t+1}\right)|b_t=0\right], \tag{68}$$

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 $V^r\left(W_{t+1},Y_t,b_t\right)$, the full rational value function. The latter would lead to the consumption CAPM holding on average: in particular, the equity premium would be too small (at $\bar{\pi}=\gamma\sigma_y^2$). Instead, the above formulation will lead to $\bar{\pi}=\gamma\sigma_r^2$, where the σ_r^2 is the volatility of the stock market.[69]

Calling $P^*\left(Y_t\right)=\frac{Y_t}{\delta}$ the price of equities in the undisturbed state, $b_t = 0$. Then, if there are no behavioral disturbances, an investor wishing to maintain a constant allocation θ^{M*} in the mixed fund should invest into it the following flow:[70]

$$\Delta F^*\left(Y^t\right)=\frac{1-\theta}{\theta\delta}\Delta Y_t \tag{69}$$

where $Y^t = (Y_t, Y_{t-1})$.

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

$$\Delta F\left(Y^t,b_t\right)=\Delta F^*\left(Y^t\right)+\Delta\left(b_tY_t\right). \tag{70}$$

Typically, we will specify that b_t is an AR(1).

The state vector is

$$Z_t=\left(Y_t,b_t\right), \tag{71}$$

where Y_t is the fundamental, and b_t the behavioral deviation.

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

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

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

[70] Indeed, the value of the mixed fund is $W^*\left(Y_t\right)=\frac{1}{\theta}P^*\left(Y_t\right)Q^{\mathcal{E}}$, and the choice of pure bond holdings consistent with θ^M is $B^*\left(Y_t\right)=(1-\theta)W^*\left(Y_t\right)$, so the flow should be $\Delta F^*\left(Y_t\right)=\frac{1-\theta}{\theta\delta}\Delta Y_t$

1. The consumer follows the consumption policy $C(Z)$, which maximizes utility subject to the above constraints.

2. The investor follows the behavioral policy (70), where the average allocation in the mixed fund is given by (68), so that it is quasi-rational with narrow framing on average, but with disturbance b_t.

3. The consumption market clears, $C(Z) = Y(Z)$.

4. The equity market clears: the mixed fund holds all the equity, so that $Q^D(Z) = Q^S$.

6.2 Model solution

Proposition 10 describes the solution of that economy. In particular, it shows that the link between the disturbance b_t and the cumulative flow $\tilde{f}_t$ is as follows. We start from an equilibrium situation, where $b_0 = 0$. The cumulative flow since time 0 can be written, to the leading order in b_t and d_t:

$$f_t = (1-\theta)\, d_t + \tilde{f}_t, \tag{72}$$

where $d_t = \sum_{s=1}^{t} \frac{\Delta Y_s}{Y_{s-1}}$ is the cumulative growth rate in the dividend, and

$$\tilde{f}_t = \theta \delta b_t. \tag{73}$$

This holds for any process b_t. This gives a microfoundation for (26).

Now, we specialize to the case where b_t follows an AR(1). Then, so does $\tilde{f}_t$, as in

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

so that we are in the "convenient benchmark" case of (26)-(27). But in this general equilibrium section, we now endogenize the interest rate and the average risk premium $\overline{\pi}$.

This AR(1) assumption is just a placeholder for richer behavioral assumptions, for example driven by time-varying beliefs (as in Caballero and Simsek (2019)), positive or negative feedback trading rules, and so on. We seek here only a coherent, simplified model, which can be fully solved and which lends itself to a number of variants.

Proposition 10. *The solution of the economy obtains in closed form as follows, taking the limit of small time intervals and only the first order terms in $\tilde{f}_t$. The market elasticity ζ and its aggregate discount factor ρ are:*

$$\zeta = 1 - \theta + \kappa^D, \qquad \rho = \frac{\zeta}{\kappa}. \tag{74}$$

The price of equities is

$$P_t = \frac{D_t}{\delta} e^{p_t}, \tag{75}$$

where $D_t = Y_t$ is the dividend, and p_t is the deviation of the price from its rational average, which depends on flows:

$$p_t = b_f^p \tilde{f}_t, \qquad b_f^p = \frac{\rho}{\rho + \phi_f} \frac{1}{\zeta}. \tag{76}$$

Hence the variance of stock market returns is

$$\sigma_r^2 = var\left(\varepsilon_t^y + b_f^p \varepsilon_t^f\right), \tag{77}$$

and hence depends on both dividend risk and flow-risk. Both affect the average equity premium, which is:

$$\bar{\pi} = \gamma \sigma_r^2, \tag{78}$$

while the risk premium at time t is affected by flows, as:

$$\pi_t = \bar{\pi} + b_f^\pi \tilde{f}_t, \qquad b_f^\pi = -\left(\delta + \phi_f\right) b_f^p. \tag{79}$$

Finally, the interest rate is constant, and given by the Euler equation (67)*:*

$$r_f = -\ln \beta + \gamma g - \gamma\left(\gamma + 1\right)\frac{\sigma_y^2}{2}, \tag{80}$$

and the average dividend-price ratio is $\delta = r_f + \bar{\pi} - g$.

Hence, we have a fairly traditional economy, except that prices are now driven by both fundamentals and flows, and the equity premium is time-varying (because of flows), and on average higher than in the CCAPM (because of narrow framing), as given in (78).

Proof sketch of Proposition 10 We provide a proof sketch here, leaving the more delicate points to the full derivation in Section 11. First, as we say, because the consumer is rational and can trade the riskless bonds, her Euler equation holds, so that (67) holds. This pins down the risk-free rate, which is constant, $R_f = \frac{1}{\beta \mathbb{E}\left[G_{t+1}^{-\gamma}\right]}$.

Next, given all net wealth is in equities (for the representative agent), the total return on wealth is $R_{t+1}^{\mathcal{E}} = \frac{P_{t+1}+Y_{t+1}}{P_t}$. Assumptions (68)-(70) imply that the financier is "rational on average", so that the Euler equation for stocks holds on average. Given that the financier maximizes a "narrow-frame" utility of wealth $V\left(W_{t+1^-}\right)$, and $W_{t+1^-} = W_t R_{t+1}^{\mathcal{E}}$, the first order condition for stock allocation is $\mathbb{E}\left[W_{t+1^-}^{-\gamma}\left(R_{t+1}^{\mathcal{E}} - R_f\right)\right] = 0$, implying $\mathbb{E}\left[\left(R_{t+1}^{\mathcal{E}}\right)^{-\gamma}\left(R_{t+1}^{\mathcal{E}} - R_f\right)\right] = 0$. Taking the limit of small time intervals, this means that the average risk premium is

$$\bar{\pi} = \gamma \sigma_r^2. \tag{81}$$

The key is the "narrow framing" with utility of wealth $V\left(W_{t+1^-}\right)$, rather than the extended utility function $V\left(W_{t+1}, Z_{t+1}\right)$, which would include a hedging demand term.

The value of the price, the risk premium, and so on, were derived in (27). □

6.3 Production: Basic equations of macro-finance with flows

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

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

Instead, the SDF (55) prices the risk-free rate and equities:

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

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

$$\begin{aligned} V\left(K_t, Z_t\right) &= \max_{I_t, L_t}\{F\left(K_t, L_t, Z_t\right)-w\left(Z_t\right)L_t-I_t-\kappa\left(I_t, K_t, Z_t\right) \\ &\quad +\mathbb{E}_t\left[\mathcal{M}_{t+1}V\left(\left(1-\delta\right)K_t+I_t, Z_{t+1}\right)\right]\} \end{aligned} \tag{82}$$

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.

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 that, 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).

This makes a number of issues that were boring, irrelevant, or inaccessible suddenly interesting, important, and researchable, thanks to the theoretical and empirical framework developed in this paper. What are the determinants of flows? Do they impact investment? How much can and should governments intervene in equity markets? Do share buybacks account for a large share of activity? What determines the low elasticity ζ of markets (perhaps contracting frictions, or plain bounded rationality), and how forward-looking are the policies of funds (κ)? Generalizing, what are the cross-market elasticities, meaning the forces that create "contagion" across market? Also, these same effects will generalize to other markets (such as the markets for corporate bonds and currencies): if so, how and what are the policy implications? This is a rich number of questions that hopefully economists will be able to answer in the coming years.

We emphasize that the "inelastic market hypothesis" remains 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.

8 Appendix: Identification methodology

8.1 Appendix: Introduction to Granular Instrumental Variables

Here we provide a brief summary of the GIV method introduced in Gabaix and Koijen (2020). Recall that we use the notations

$$X_E := \frac{1}{N}\sum_{i=1}^{N} X_i, \qquad X_S := \sum_{i=1}^{N} S_i X_i, \qquad X_\Gamma := X_S - X_E. \tag{83}$$

We also call

$$\tilde{E}_i = \frac{\tilde{\sigma}_i^{-2}}{\sum_{j=1}^{N}\tilde{\sigma}_j^{-2}}$$

the "quasi-equal weight", and $X_{\tilde{E}} := \sum_{i=1}^{N}\tilde{E}_i X_i$ the quasi-equal weighted average.

Suppose that we have a time series of changes in investors' equity holdings, Δq_{it}, that can be modeled as:

$$\Delta q_{it} = -\zeta \Delta p_t + \lambda_i \eta_t + u_{it}, \tag{84}$$

where ζ is the demand elasticity of interest, Δp_t the price change, η_t is a common shock (or a vector of those), λ_i a factor loading, and u_{it} an idiosyncratic shock. The GIV method identifies ζ using variation that comes from the idiosyncratic shocks, u_{it}.

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

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

for the multiplier

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

The goal is to estimate M, which identifies the demand elasticity, ζ. We provide here a summary of the method, which extends to much richer cases.

Simple case with uniform loadings We start with the case where $\lambda_i = 1$, so that all loadings on the common shocks are uniform. Then, we form the GIV:

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

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:

$$z_t = \left(-\zeta \Delta p_t + \eta_t + u_{St}\right) - \left(-\zeta \Delta p_t + \eta_t + u_{Et}\right) = u_{St} - u_{Et},$$

meaning that

$$z_t = u_{\Gamma t} := u_{St} - u_{Et}.$$

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

$$\Delta p_t = M z_t + e_t,$$

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

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

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

$$\Delta \check{q}_{it} = \sum_{f=1}^{r} \check{\lambda}_i^f \eta_t^f + \check{u}_{it}. \tag{85}$$

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

$$\Delta p_t = M z_t + \sum_f \beta^f \eta_t^f + e_t, \tag{86}$$

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

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

8.2 Specific algorithms

8.2.1 Algorithm used for sector-level data

In this section, we specify the steps that we take to construct the instrument used with the Flow of Funds (FoF) data in Section 4.4, which boils down to recovering u_{it}.

1. We construct $\tilde{E}_{j,t-1}$ weights, where we start from quasi-equal weights, where $\sigma_j = \sigma(\Delta q_{jt})$,

$$\tilde{E}_j^{\sigma} = \frac{\sigma_j^{-2}}{\sum_{k=1}^{N} \sigma_k^{-2}},$$

   and define $\tilde{E}_j = \min\left\{\tilde{E}_j^{\sigma}, \frac{1.5}{N}\right\}$, which limits 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_{jt}. This is particularly relevant when the number of sectors is small, as is the case for the FoF.

2. We run the panel regression

$$\Delta q_{jt} = \alpha_j + \beta_t + \gamma_j \Delta y_t + \delta_j t + \Delta \check{q}_{jt}, \tag{87}$$

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

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

4. We construct the GIV instrument[71]

$$Z_t = \sum_{j=1}^{N} S_{j,t-1}\Delta\check{q}_{it}, \tag{88}$$

excluding the corporate sector.

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

$$\Delta p_t = \alpha + MZ_t + \lambda_P'\eta_t^e + \beta_P\Delta y_t + e_t. \tag{89}$$

This regression is also the first stage to estimate the elasticities. We estimate the demand elasticity using

$$\Delta q_{Et} = \alpha_E - \zeta\Delta p_t + \lambda_E'\eta_t + \beta_E\Delta y_t + e_t, \tag{90}$$

and the supply elasticity via

$$\Delta q_{Ct} = \alpha_C - \zeta_C\Delta p_t + \lambda_C'\eta_t + \beta_C\Delta y_t + e_t, \tag{91}$$

where we instrument Δp_t using Z_t in both cases.

8.2.2 Algorithm used for investor-level data

There are two potential advantages of using 13F data over FoF data. First, we have a much larger cross-section of investors, which allows us to estimate the factors, η_t^f, more accurately. Second, we can construct investor characteristics, $x_{j,t-1}$, which allows us to estimate parametric factors as well. We discuss the selection of the sample and construction of the characteristics in Section 9.3.

The algorithm that we use in Section 4.5 is given by the following steps.

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

2. We run the panel regression

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

using $\frac{1}{\sigma_j^2}$, with $\sigma_j = \sigma(\Delta q_{jt})$,[72] as regression weights, Δy_t being quarterly real GDP growth, and I_t denotes institutional ownership with captures the time trend as in (87).[73]

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

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

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

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

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

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

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

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

4. We construct the instrument, $z_t^* = \sum_j S_{j,t-1}^{\star}\check{u}_{jt}$, where $S_{j,t-1}^{\star}$ are the size weights that add to one for the institutions used in the estimation. We also define κ_{t-1} to be the share of the aggregate stock market held by these institutions. We define $z_t = \kappa_{t-1}z_t^{\star}$.

5. We estimate the multiplier via the regression

$$\Delta p_t = \alpha + Mz_t + \gamma\Delta y_t + \epsilon_t.$$

References

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

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.

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

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

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

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

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

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.

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

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

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

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

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

Campbell, John Y. and Robert J. Shiller, "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, 1988, *43* (3), 661–676.

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

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

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

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

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

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

Di Maggio, Marco, Amir Kermani, and Kaveh Majlesi, "Stock market returns and consumption," *Available at SSRN 3103538*, 2018.

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

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.

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, "A Behavioral New Keynesian Model," *Forthcoming at the American Economic Review*, 2020.

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

Mitchell, Mark, Lasse Heje Pedersen, and Todd Pulvino, "Slow moving capital," *American Economic Review*, 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.

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

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.

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.

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

July 8, 2020

9 Appendix: Data Sources and Construction

9.1 Data sources

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

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

Exchange-traded funds (ETFs) We use data on ETF flows, assets under management, and the share invested in U.S. equities from Morningstar. The sample is quarterly from 1993.Q1 to 2017.Q2.

State and local pension funds We also decompose the holdings of state and local pension funds to assess the heterogeneity in the bond-stock allocation, which is a key input into our model. The sample is from 2012.Q1 to 2017.Q1.[74]

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

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

Asset prices Data on returns with and without dividends are from the Center for Research in Security Prices at the University of Chicago, Booth School of Business. We use the monthly, value-weighted return with and without dividends to compute the monthly dividend payment. To compute dividend growth, we sum the dividends during a 12-month period and we compute the geometric annual growth rate.

[74] The structure of the reporting requirements changed in 2012, which motivates the start of our sample. However, we use these data only to illustrate the heterogeneity across funds, for which the shorter sample suffices.

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

9.2 Sector-level data: Flow of Funds

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

9.2.1 Data items

We use Corporate Equities (Table 223) for equities and for fixed income we take the sum of Treasury Securities (Table 210) and Corporate and Foreign Bonds (Table 213).

We use unadjusted flows (FU) and, for the levels, we use the unadjusted market values when available (LM) and otherwise the estimated level (FL). Net issuances are initially equal to the total aggregate flow. The Flow of Funds revises historical data every quarter.

9.2.2 Notation and data definitions

Sectors are indexed by $i = 1, ..., I$, where $i = Foreign$ refers to the foreign sector. We observe holdings of equities, $\mathcal{E}_{it}$, Treasuries, Tr_{it}, and corporate bonds, C_{it}. We refer to the sum of Treasuries and corporate bonds as bonds, $B_{it} = Tr_{it} + C_{it}$. The flows corresponding to each asset class are denoted by ΔF_{it}^a, $a = \mathcal{E}, Tr, C, B$. Aggregate levels and flows omit the subscript i, implying, for instance, for equities $\sum_i \mathcal{E}_{it} = \mathcal{E}_t$ and for bonds $B_t = \sum_i B_{it}$. Lastly, the gross capital gain for equities is denoted by R_t^X and the return inclusive of dividend payments is denoted by R_t.

We define the total assets of sector i as $W_{it} = \mathcal{E}_{it} + B_{it}$. Net issuances, $ni_t = \frac{NI_t}{\mathcal{E}_{t-1}}$, are based on equity markets.

In the FoF, equity flows are defined by $\Delta F_{it}^{\mathcal{E}} = \mathcal{E}_{it} - \mathcal{E}_{i,t-1} R_t^X$.[75] We assume in what follows that the securities are adjusted at the end of the quarter, $\Delta F_{it}^a = \Delta F_{it}^a = \left(\Delta Q_{it}^a\right) P_t^a, a = \mathcal{E}, B$. The total per-period flow is $\Delta F_{it} = \Delta F_{it}^{\mathcal{E}} + \Delta F_{it}^B$ and in relative terms $\Delta f_{it}^a = \frac{\Delta F_{it}^a}{\mathcal{E}_{i,t-1}}, a = \mathcal{E}, B$. In the model, the cumulative flow is defined as

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

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

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

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

9.2.3 Adjustment for foreign holdings of equity and corporate bonds

The FoF reports total flows and holdings of corporate equities and corporate bonds, including foreign assets held by US investors. As we are interested in measuring the flow into the US equity market, we adjust the holdings and flows for foreign positions. Unfortunately, we do not know the holdings and flows of foreign assets by sector, but we do know the aggregate positions across investors. We discuss our measurement approach in the context of equities, but we apply the same procedure to corporate bonds.[76]

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

We start from the following identities

$$x_{it}^{D} + x_{it}^{F} = x_{it}^{T}, \tag{92}$$

$$\mathcal{E}_{it}^{j} = \mathcal{E}_{i,t-1}^{j} R_t^{X,j} + \Delta F_{it}^{j}, \tag{93}$$

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

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

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

It then follows that

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

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

For flows, we assume that

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

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

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

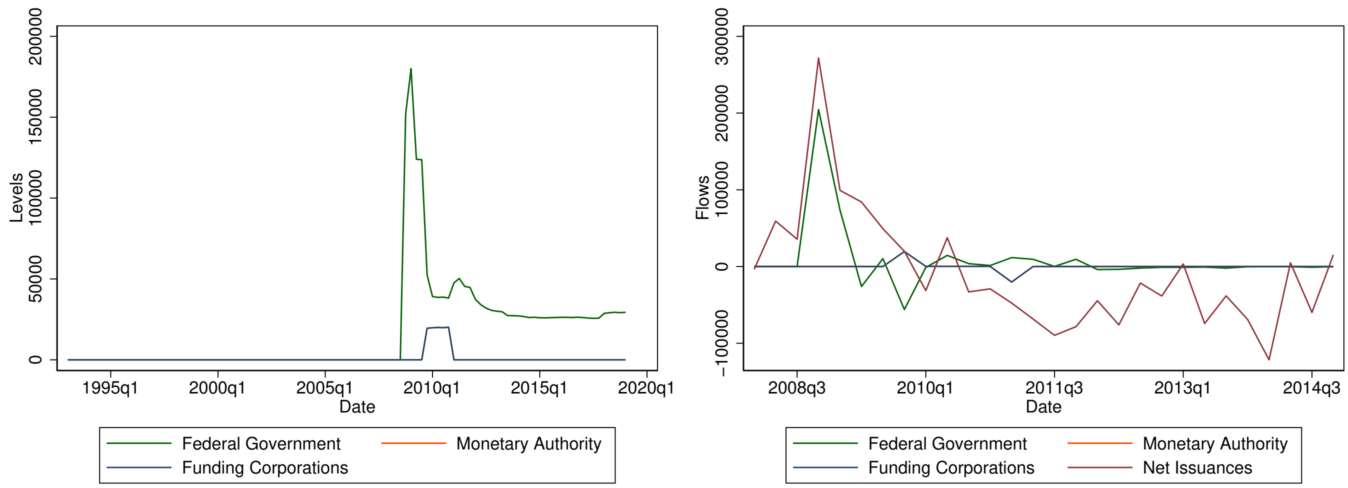
implying

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

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

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

Figure 8: 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.



9.2.4 The impact of the 2008-2009 financial crisis

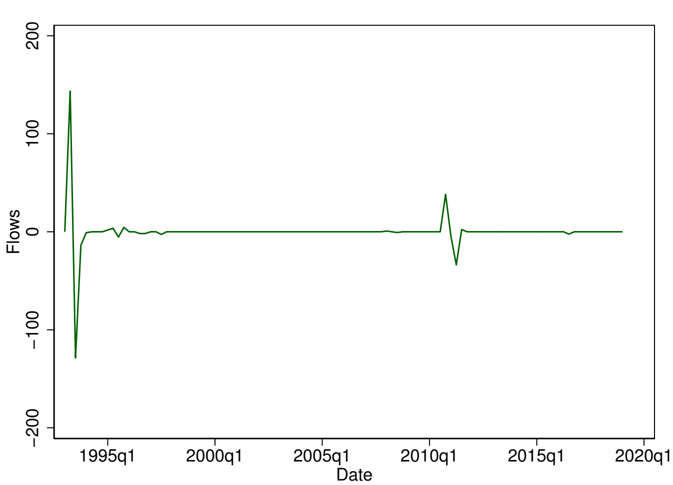
Three sectors have non-zero equity holdings only since the 2008-2009 financial crisis: the Federal Government (sector 31), the Monetary Authority (sector 50), and Funding Corporations (sector 71). These positions are all associated with the federal financial stabilization programs. We describe the adjustments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

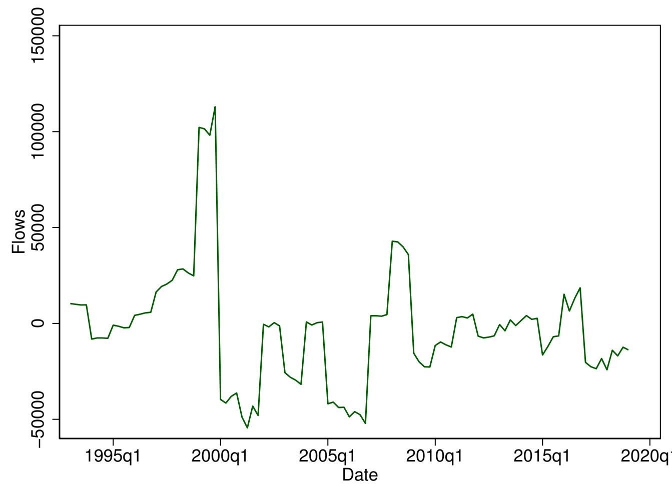
The holdings of the Federal Government are derived from "corporate equities issued by commercial banking under the federal financial stabilization programs," "corporate equities issued by funding corporations (AIG) under the federal financial stabilization programs," "corporate equities issued by bank-holding companies (GMAC) under the federal financial stabilization programs," 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 8 from 1993.Q1 to 2018.Q4. The dynamics of net issuances alongside the flows associated with the three sector are plotted in the right panel of Figure 8 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.

Figure 9: Flows of Foreign Banking Offices in the U.S. and Non-financial Corporate Businesses





9.2.5 Foreign banking offices in the U.S. and non-financial corporate business holdings

We make adjustments for two additional sectors. First, the sector Foreign Banking Offices in the U.S. (sector 75) has largely zero holdings since 1993, see the left panel of Figure 9. Second, for the asset holdings of non-financial corporate businesses (sector 10), the quarterly flows are poorly measured, see the right panel of Figure 9 showing the series from 1993.Q1 to 2018.Q4. The reason is that the FoF interpolates annual flows.[77] These flows and holdings reflect firms' holdings of other firms' equity, for instance for strategic or speculative reasons. Prior to the September 2018 publication, the FoF showed the equity liability of the non-financial corporate sector net of these 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.

9.2.6 Examples of measurement issues

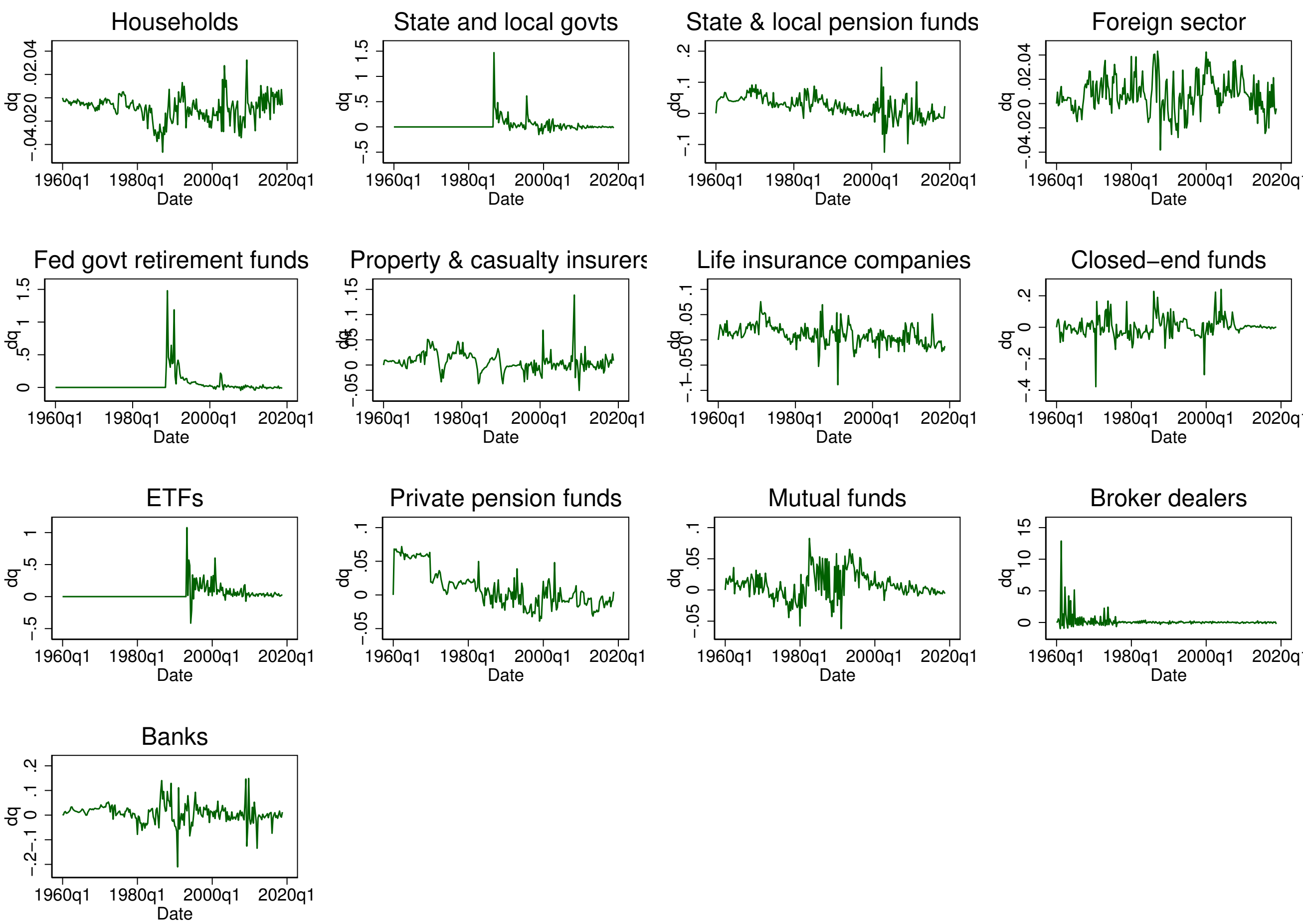
Even though the FoF data are the best data to use for both equity and fixed income holdings, certain measurement issues remain. We list them here and perhaps future research can refine some of our calculations. First, in the FoF, shares issued by ETFs, closed-end funds, and real estate investment trusts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

Similarly, the total holdings include closely held equity. While the supply side is separated, we do not have disaggregated holdings, which implies we cannot adjust for this on the demand side.

We start our sample in the early nineties. In part, this starting date is driven by the fact that institutional ownership has been rising, which allows us to provide a better disaggregation of the different sectors. However, the dynamics of equity flows, Δq, also looks more erratic in the earlier years. In Figure 10, we plot the dynamics of equity flows across sectors to illustrate this issue. We therefore start in 1993. Lastly, ETFs become available in 1993. ETFs have been growing since then, in part to replace mutual funds, and we merge ETFs and mutual funds for parts of our analysis.

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

Figure 10: Dynamics of equity flows across sectors. The figure shows the equity flows for the final 13 sectors in our sample from 1960.Q1 to 2018.Q4.



9.2.7 Sample construction for the GIV estimator

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

Second, we winsorize the data by first removing the time-series median of each series, which is a robust way to remove differences in the levels of the series. We then winsorize each series across time and sectors for the period from 1993.Q1 to 2006.Q4 to mitigate the influence of outliers. This avoids the need to winsorize during the financial crisis and the larger, as well as in the case of the more volatile flows happening during the earlier part of the sample. Winsorizing the data unconditionally does not impact our results much.

9.3 Investor-level data: 13F filings

We use 13F data from FactSet, which has also been used in Koijen et al. (2019).

9.3.1 Measuring changes in equity demand

We first discuss how we construct flows. 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 $\mathcal{E} = \sum_a H^-_{iat}$, where $H^-_{iat} = \frac{H_{iat}}{1+R^X_{at}}$, where R^X_{at} is the capital gain. In the absence of (reverse) splits, it holds $H^-_{iat} = Q_{iat}P_{a,t-1}$. We now define the flow as

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

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

9.3.2 Constructing characteristics

Next, we discuss how we construct the characteristics that we use in Section 8.2.2. We define the following characteristics

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

2. Active share, which is defined as

$$\frac{1}{2}\sum_a \left|\theta_{iat} - \theta^m_{iat}\right|,$$

where θ_{iat} is the portfolio share and θ^m_{iat} the market-weighted portfolio of securities held by investor i at time t. We compute the eight-quarter median to mitigate the impact of outliers.

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

4. Portfolio turnover, which is defined as

$$\sum_a \frac{\left|H_{iat}^{-} - H_{ia,t-1}\right|}{\mathcal{E}_{it}^{*}}.$$

   As turnover is skewed, we use the log of 0.01 plus turnover as the characteristic. We the compute the eight-quarter median to mitigate the impact of outliers.

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

9.3.3 Sample selection

We include an investor's flow, Δq_{it}, when we observe the flow and the lagged characteristics, $x_{i,t-1}$. Second, we remove investor-quarter observations where the flow in period t exceeds 100% and is below -100% in the next quarter (and vice versa). Such cases, although rare, may be data errors. We drop investor-quarter observations in which investors hold fewer than 25 stocks or invest less than $50 million in equities. After these screens, we only include investors that satisfy these screens for at least 25 quarters. We impose the latter screen to ensure that we can estimate the factor loadings and non-parametric loadings sufficiently accurately. In the final sample, we omit the household sector, which is measured as a residual.[78]

We winsorize flows and investor characteristics, each period, at the 2.5%- and 97.5%-percentiles, with the exception of size.

10 Additional Figures and Tables

10.1 Drawdown dynamics

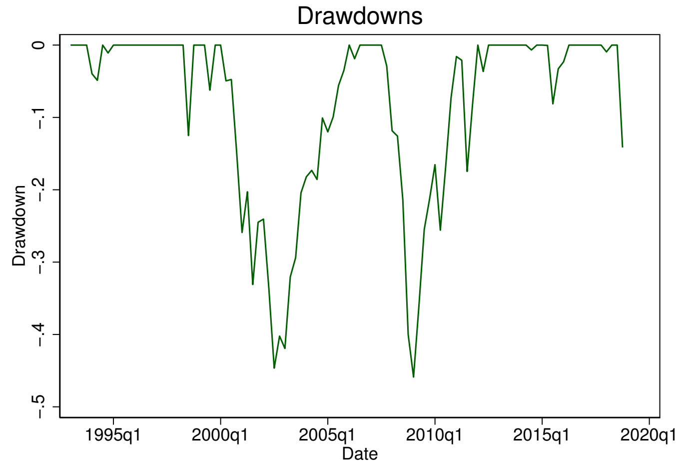
In Figure 11, we plot the drawdowns, defined as the decline in the cumulative stock market index relative to its maximum so far, of the CRSP value-weighted index. We use these drawdowns to date recessions that we study in Section 2.

10.2 Flows between sectors

To assess the magnitude of equity risk reallocation across sectors, we compute $y_t^{Gross} = \frac{\sum\left|\Delta F_{jt}^{\mathcal{E}}\right| + \left|\Delta F_t^{Firm}\right|}{2\mathcal{E}_{t-1}}$, where ΔF_t^{Firm} denotes net issuances of equity by firms. We divide the measure by two as for every

[78]However, this also means that all measurement errors end up in the household sector. In addition, in the presence of short-selling activity, this also distorts the holdings of the household sector.

Figure 11: The figure illustrates the drawdowns of the U.S. 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.



buyer of \$1 of equity, there is a seller of the same amount. As some of the flows are associated with net repurchases, we separately measure the equity risk "creation" and "redemption" as a result of such corporate actions via $y_t^{AbsNet} = \frac{|\Delta F_t^{Firm}|}{\mathcal{E}_{t-1}}$, which we will refer to as absolute net flows.

In Figure 12, we plot the average flows between sectors from 1993.Q1 to 2018.Q4.

Figure 12: 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.

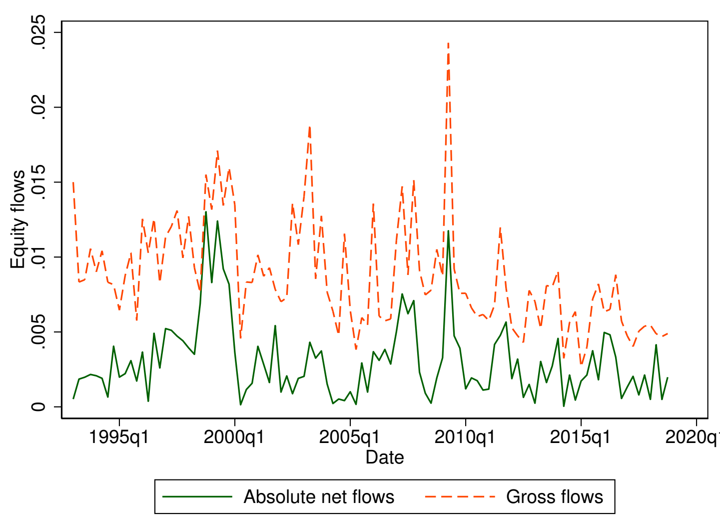
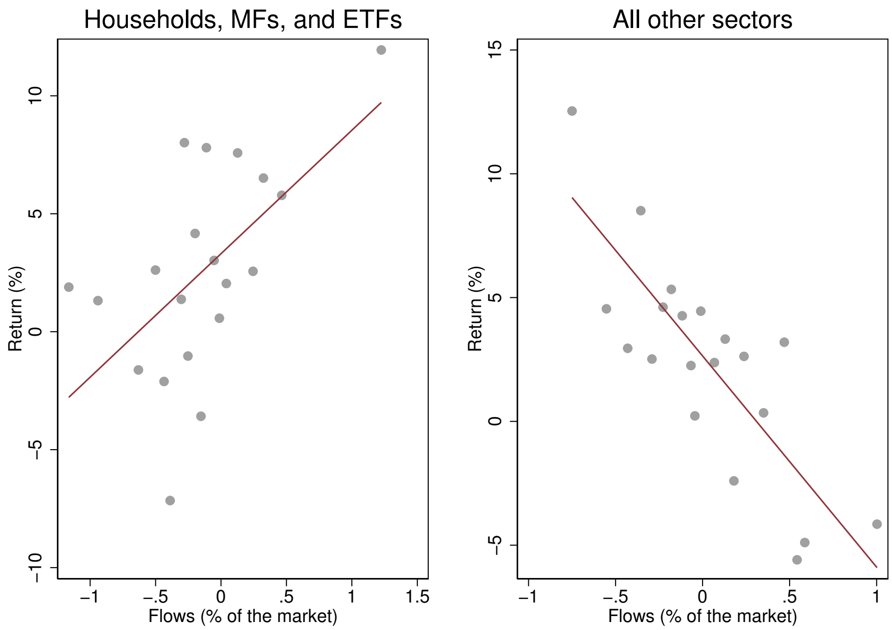


Figure 13: The figure illustrates the connection between flows of the household sector (Panel A) and all other sectors (Panel B) during the recent sample from 2000 to 2018.



10.3 Preliminary evidence on the link between equity flows and prices and inter-sector flows

Motivated by the case studies in Section 2, we explore a preliminary potential implication for prices. We emphasize that this is just a preliminary result, and Section 4 explores the link between prices, quantities, and flows in much greater detail. Indeed, suppose that markets move, and perhaps in particular during bad times, due to demand shocks and flows coming from households (and other investors that are included in the Flow of Fund's definition of the household sector), and that those demand shock are accommodated by other investors who are fairly inelastic. If this is the case, we would expect to see a positive correlation between the equity flows coming from households and returns, and a negative correlation between the rebalancing of all other sectors, such as state and local pension funds and foreign investors, and returns. As a large fraction of households' investments is now done via ETFs and mutual funds, we group them together with households. If, on the other hand, common demand shocks drive prices, or demand shocks from different investor types matter at different points in time, we would not detect a correlation between flows and prices.

Figure 13 provides a binned scatter plot of the two groups with flows (again scaled by the size of the aggregate stock market) against returns. The volatility of the flow series is 42bp and 51bp per quarter, respectively, showing that we are capturing a large fraction of the risk being reallocated in equity markets (see Figure 12). Consistent with the experience during the two recessions, shocks to households and related sectors are positively correlated to prices, while the correlation between the flows of all other sectors and returns is negative. In the extreme case in which the demand shocks only come from otherwise inelastic households, mutual funds, and ETFs, while the other sectors provide elasticity but do not experience demand shocks themselves, the (negative of the) slope in Panel B of Figure 13 measures the macro demand elasticity. The slope would correspond to a demand elasticity of approximately 0.1. Put differently, a 1% demand shock would result in a 10% change in prices. All this is only suggestive, and we develop a more rigorous estimate of the demand elasticity in Section 4.4.

10.4 Robustness of the GIV estimates from the FoF

In this section, we summarize the results from different measurement assumptions to estimate the multiplier, M. In Table 4, we first repeat the benchmark estimate as a point of reference in the first column. In the second column, we omit all principal components. In the third column, we do not merge mutual funds and ETFs. In the fourth column, we do merge mutual funds and ETFs again, but now winsorize over the full sample from 1993.Q1 to 2018.Q4 instead of from 1993.Q1 to 2006.Q4. In column five, we omit the time trend in (87). In column six, we control for the lagged value of Δq_{jt} in (87), allowing for heterogeneous persistence coefficients across sectors. Lastly, in the seventh column, we start the sample in 2000.Q1.

The main takeaway is that the multiplier estimates are quite stable across the various specifications. The estimates of the multiplier vary between 6.6 and 8.0.

Table 4: 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 (87), allowing for heterogeneous persistence coefficients across sectors. Lastly, in the seventh column, we start the sample in 2000.Q1. Standard errors are reported in parentheses.

	Δp	Δp	Δp	Δp	Δp	Δp	Δp
Z	7.08	8.00	6.94	7.65	6.63	6.77	6.79
	(1.86)	(1.24)	(1.48)	(1.37)	(1.17)	(2.18)	(2.02)
GDP growth	5.99	5.99	6.06	6.02	6.14	6.01	6.20
	(0.69)	(0.59)	(0.66)	(0.75)	(0.69)	(0.73)	(0.74)
η_1	21.06		22.09	15.59	32.54	25.24	-30.65
	(13.58)		(11.11)	(11.58)	(6.40)	(13.64)	(15.65)
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)
Observations	104	104	104	104	104	103	76

In Table 5, we replicate Table 1, but now adding the lagged value of Z_t. Across various specifications, Z has a small positive autocorrelation of approximately 10-15%. As is clear from comparing both sets of estimates, including a lag does not change the estimates in a meaningful way, and the lagged value of Z is in all cases insignificant.

In Table 6, 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

Table 5: Robustness of the GIV estimates: Persistence in Z. The table replicates Table 1, but now adding the lagged value of Z_t. The sample is from 1993.Q1 to 2018.Q4. Standard errors are reported in parentheses.

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

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 6. 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.

10.5 The volatility of idiosyncratic demand shocks and measuring flows

In Table 7, we report the standard deviation of sector-specific demand shocks, u_{it}, as constructed and used in Section 4.6.

In Table 8, we report the slope estimates, β_j, of the regression in equation (49). 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 6: Robustness of the GIV estimates: Additional principal components.The table adds principal components starting from a single principal component. The sample is from 1993.Q1 to 2018.Q4. Standard errors are reported in parentheses.

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

Table 7: Volatility of idiosyncratic demand shocks by sector. The table reports the volatility of idiosyncratic demand shocks by sector. We follow the procedure outlined in Section 4.6 to estimate the demand shocks. The sample is from 1993.Q1 to 2018.Q4.

Sector	$S\sigma(u_{it})$	$\sigma(u_{it})$
Households	0.61	1.51
Mutual funds	0.21	0.90
Foreign sector	0.19	1.48
State & local pension funds	0.19	2.53
Private pension funds	0.12	1.19
Broker dealers	0.032	6.40
Life insurers	0.026	1.42
State and local govts	0.019	3.48
Property & casualty insurers	0.017	1.48
Closed-end funds	0.012	3.20
Fed govt retirement funds	0.0092	2.08
Banks	0.0092	3.433

Table 8: Assessing the mismeasurement of capital flows. The table reports the slope coefficient of the regression in equation (49) to assess whether capital flows are mismeasured. We consider a flow to be correctly measured when $|\beta_j - 1| < 0.3$. The sample is from 1993.Q1 to 2018.Q4.

Sector	Flows included	b_f	T-statistics for $H_0 : b_f = 1$
Households	0.00	0.47	9.11
State and local govts	1.00	0.49	1.95
State & local pension funds	1.00	1.04	0.73
Foreign sector	0.00	0.35	6.90
Fed govt retirement funds	0.00	-0.03	12.08
Property & casualty insurers	0.00	0.55	3.30
Life insurance companies	0.00	0.61	2.61
Closed-end funds	1.00	1.08	0.85
ETFs	1.00	1.01	0.93
Private pension funds	1.00	1.10	1.49
Mutual funds	1.00	0.98	0.53
Broker dealers	0.00	0.07	22.78
Banks	0.00	-0.06	10.67

11 Appendix: Omitted Proofs

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

$$d\theta = \frac{d\pi}{\pi}. \tag{94}$$

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

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

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

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

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

$$\frac{\Delta Q}{Q^{\mathcal{E}}} = \frac{\Delta\pi}{\psi\pi}. \tag{95}$$

If the value of equity changes by p, the risk premium changes by $\Delta\pi = -\delta p$ (see (17)), so we have

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

where the rational elasticity is:

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

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

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

which is the announced expression.

Case 2: Disaster risk. The reasoning is the same, except that expression (94) is different with disaster risk. To derive it, observe that the value function must take the form $V(W_t) = KW_t^{1-\gamma}$ for some constant K. Hence, calling $\tilde{R}_{t+1}$ the rate of return on stocks, the consumer's problem

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

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

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

so the FOC characterizing the equity share is:

$$\mathbb{E}\left[\left(1+\theta\tilde{r}_{t+1}\right)^{-\gamma}\tilde{r}_{t+1}\right] = 0. \tag{96}$$

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 risk premium conditional on no disasters, where $J_t = 0$ if there is no disaster and 1 otherwise. Then (96) becomes

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

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

$$\pi = p\left(1-\theta(1-B)\right)^{-\gamma}(1-B). \tag{97}$$

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

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

i.e.

$$d\theta = \frac{d\pi}{\pi}\frac{B}{\gamma(1-B)}, \tag{98}$$

which is the disaster counterpart to (94): how the desired equity share changes as the risk premium changes.

The rest of the derivation is exactly as in the lognormal Case 1, replacing (94) by (98).

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 risk premium π by the perceived risk premium).

Proof of Proposition 8 Calling $\mathcal{D}_1$ the aggregate dividend, the dividend per share goes from $D_1 = \frac{\mathcal{D}_1}{Q_0}$ to $D_1' = \frac{\mathcal{D}_1}{Q_0'} = \frac{D_1}{1-b}$. So, the time-1 dividend per share increases by $d = b$.

Let us first consider a frictionless, elastic / rational model. The price per share increase by the same fraction as the time-1 dividend per share, i.e. $p = b$. Calling $v = \Delta\ln(PQ) = p + q^S$ the change in the market value of the firm, we have:

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

The market value does not change: the lowering of the number of share outstanding by b is compensated by the increase in the price per share by a fraction, which is the same b.

Let us next consider an inelastic model. The buyback decreases the total dividend payout from D_0 to $D_0 - P_0 Q_0 b$, so "lowers the dividend from corporates to the fund by $f^{C\to M} = \frac{-P_0 Q_0 b}{W} = \frac{-\theta W b}{W} = -\theta b$, so

$$f^{C\to M} = -\theta b.$$

Calling f^h the flow from households, the total flow is $f = f^h + f^{C\to M} = f^h - \theta b$. Hence, the "supply equal demand" condition in the share market translates into:

$$q^S = -b = q^D = -\zeta p + \kappa \delta d + f^h - \theta b,$$

so $\zeta p = \zeta b + f^h$ and

$$\text{Price change after buyback:} \quad p = b + \frac{f^h}{\zeta}. \tag{99}$$

We see that in the "frictionless limit" $\zeta \to \infty$, $p = b$, like in the rational model. But otherwise, it depends on f^h, the household flows.

We explore some modelling of f^h. Call μ^D (respectively μ^G) the fraction of the dividend (respectively, of the capital gain) that is "absorbed" by the households, that is, consumed, or reinvested in bonds. Then, the primitive flow is $f^{h,0} = -\mu^D \theta \frac{\mathcal{D}_0'}{P_0 Q_0} - \mu^G \theta p$, so that the change in the flow is caused by the buyback is:

$$f^h = \mu^D \theta b - \mu^G \theta p.$$

For instance, the capital gain per share is p, the capital gain as a fraction of funds' assets is θp, and a fraction μ^G of that is "absorbed" (removed from the funds) by households. Hence, (99) gives:

$$p = \frac{\zeta + \mu^D \theta}{\zeta + \mu^G \theta} b. \tag{100}$$

This yields (50) and implies that $p > b$ if $\mu^D > \mu^G$. A share buyback increases the market value by $v = p + q^S = p - b > 0$.

Proof of Proposition 10 First, we derive the risk-free rate. The consumer's first order condition gives the Euler equation $1 = \beta R_{f,t} \mathbb{E}_t \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \right]$. As in equilibrium $C_t = Y_t$, the interest rate satisfies the Euler equation for bonds.

Now, we move to stocks. The average allocation in equities maximizes a risk-adjusted return, $\mathbb{E}_t \left[V^p \left(R_{t+1} \right) \right]$, with $V^p(R) = \frac{R^{1-\gamma}-1}{1-\gamma}$. Then, approximately, the allocation in equities is $\bar{\theta}^{\mathcal{E}} = \frac{\bar{\pi}}{\gamma \sigma_r^2}$. Given that in equilibrium all the wealth comes from equity, the risk premium is $\bar{\pi} = \gamma \sigma_r^2$. The rest is derived in the paper.

Consider the economy when the behavioral disturbance b_t is at 0. Then, the price of equities is $P_t = \frac{Y_t}{\delta}$. So, the mixed fund holds a quantity of bonds

$$B_t^M = \frac{1-\theta}{\theta} Q^{\mathcal{E}} P_t = \frac{1-\theta}{\theta \delta} \frac{Y_t}{\delta}.$$

Hence the rational flow in the pure bond fund is

$$\Delta F_t^r = \psi \Delta B_t^r, \qquad \psi = \frac{1-\theta}{\theta \delta}$$

With the disturbance b_t, the flow is

$$\Delta F_t = \psi \Delta Y_t + \Delta (b_t Y_t),$$

i.e., calling $W_t^* = \frac{1}{\theta\delta} Y_t$ the size of the mixed fund if there is not friction.

$$\Delta f_t = \frac{\Delta F_t}{W_{t-1}^*} = \frac{\psi \Delta Y_t}{B_Y^W Y_{t-1}} + \frac{\Delta (b_t Y_t)}{Y_{t-1}} = (1-\theta) \frac{\Delta Y_t}{Y_{t-1}} + \theta\delta \frac{\Delta (b_t Y_t)}{Y_{t-1}}$$

hence

$$f_t = (1-\theta) d_t + \tilde{f}_t, \tag{101}$$

where, to the leading order

$$\tilde{f}_t = \theta\delta b_t,$$

and

$$d_t = \sum_{s=1}^{t} \frac{\Delta Y_s}{Y_{s-1}}$$

is the cumulative increase in the dividend, which is GDP.

The expressions for price and the equity premium were derived in (27). Note that what we call p_t here the value of $p_t - d_t$ in (27).

12 Appendix: Theory Complements

12.1 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

$$\pi_t = \mathbb{E}_t \left[\frac{dP_t}{P_t}\right] / dt + \frac{D_t}{P_t} - r \tag{102}$$

has the Taylor expansion:

$$\hat{\pi}_t = \mathbb{E}_t \frac{dp_t}{dt} + \delta (d_t - p_t). \tag{103}$$

We have

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

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

$$\mathbb{E}_t \left[\frac{dp_t}{dt}\right] - \rho p_t + \delta d_t + \frac{\nu_t + f_t}{\kappa} = 0. \tag{104}$$

Integrating forward, the stock price is:

$$p_t = \mathbb{E}_t \int_t^\infty e^{-\rho(\tau - t)} \left(\frac{\rho}{\zeta} (f_\tau + \nu_\tau) + \delta d_\tau\right) d\tau \tag{105}$$

$$= \frac{\delta}{\rho} d_t + \frac{1}{\zeta} (f_t + \nu_t) + \mathbb{E}_t \int_t^\infty e^{-\rho(\tau - t)} \left(\frac{\dot{f}_\tau + \dot{\nu}_\tau}{\zeta} + \frac{\delta}{\rho} \dot{d}_\tau\right) d\tau. \tag{106}$$

This allows for easier calculations than the discrete time model. For instance, suppose that flows and dividends follow autoregressive process, e.g. $df_t = -\phi_f f_t + \sigma_t dZ_t$ (for dZ_t a mean-zero increment process, e.g. a Brownian motion). Then we have $\mathbb{E}_t [f_\tau] = e^{-\phi_f(\tau-t)} f_t$ for $\tau \geq t$, so (105) gives:

$$p_t = \frac{\rho}{\rho + \phi_f} \frac{f_t}{\zeta} + \frac{\delta}{\rho + \phi_d} d_t. \tag{107}$$

In the random walk case ($\phi_f = \phi_d = 0$), we recover the formula seen in the one-shot case (4), where we had $\delta = 1$. As always, more persistent shocks (i.e. lower ϕ shocks) have a higher impact.

Combining (107) and (22) leads to:

$$\hat{\pi}_t = b_f^\pi f_t + b_d^\pi d_t, \tag{108}$$

for two coefficients $b_f^\pi < 0$ 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 $b_d^\pi = \frac{\delta(1-\theta)}{\zeta+\kappa\phi_d}$.

12.2 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:

$$p_t = \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho}{(1+\rho)^{\tau-t+1}} \frac{f_\tau + \nu_\tau}{\zeta_\tau} \tag{109}$$

To be concrete, we study the case

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

so that $\mathcal{M}_t$ is a temporary increase in market inelasticity, mean-reverting at a speed ϕ_ζ. We consider the impact of a permanent inflow, $f_\tau = f_0$ for $\tau \geq 0$. Then the price follows:

$$p_t = \frac{f_0}{\zeta}\left(1 + \frac{\rho}{\rho + \phi_\zeta}\mathcal{M}_t\right). \tag{110}$$

So, if the flow f_0 happens during a time of high market inelasticity $\frac{1}{\zeta_t}$ (i.e. high $\mathcal{M}_t$), then the price impact is higher, which makes sense. It is the average future value of the inelasticity $(\frac{\rho}{\rho+\phi_\zeta}\mathcal{M}_0)$ that matters, rather than the current inelasticity ($\mathcal{M}_0$). In the scenario above, the price impact f_0 mean-reverts at a speed ψ.

More generally (if $\phi_\zeta > \phi_f$), this implies implies that returns that happened during a high-volatility period mean-revert faster.

A tentative calibration. With $\rho = 0.13/year$ and $\psi = 0.15/year$, we have $\frac{\rho}{\rho+\psi} \simeq 0.4$, so have then to get a price impact higher by a factor 0.5, we need $\mathcal{M}_t = \frac{0.5}{0.4} = 1.2$, i.e. a halving of ζ_t.

12.3 When flows react to the risk premia

Here we derive Proposition 9, and more generally explore the consequences of flows of the type:

$$\Delta f_t = \chi \hat{\pi}_t + \varepsilon_t. \tag{111}$$

We first proceed in continuous time, which is cleanest.

12.3.1 Continuous time

For simplicity, we assume away dividend surprises. They would be easy to add back. The flows (111) are

$$df_t = \chi \hat{\pi}_t dt + \sigma dz_t. \tag{112}$$

We use the operator D,

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

So, $\hat{\pi}_t = -\delta p_t + \frac{\mathbb{E}_t[dp_t]}{dt}$ (see Section 12.1) becomes:

$$\hat{\pi}_t = (D - \delta) p_t \tag{113}$$

and (112) gives

$$Df_t = \chi \hat{\pi}_t. \tag{114}$$

The basic dynamic pricing equation, (23), becomes:

$$0 = -\zeta p_t + \kappa D p_t + f_t. \tag{115}$$

Differentiating once and taking time-t expectations gives:

$$\begin{aligned} 0 &= -\zeta D p_t + \kappa D^2 p_t + D f_t & (116) \\ &= \left[-\zeta D + \kappa D^2 + \chi (D - \delta)\right] p_t & (117) \\ &= H(D) p_t \end{aligned}$$

where

$$H(x) = \kappa x^2 - (\zeta - \chi) x - \chi \delta. \tag{118}$$

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:

$$\rho = \frac{\zeta - \chi + \sqrt{\Delta}}{2\kappa}, \qquad \phi = \frac{-\zeta + \chi + \sqrt{\Delta}}{2\kappa}, \qquad \Delta = (\zeta - \chi)^2 + 4\chi\kappa\delta. \tag{119}$$

When $\chi = 0$, $\rho = \frac{\zeta}{\kappa}$ (as in Proposition 5) and $\phi = 0$. We record that ϕ solves:

$$(\zeta + \kappa\phi) \phi = \chi (\phi + \delta) \tag{120}$$

and as the product of the two roots, $-\phi\rho$ is equal to $\frac{-\chi\delta}{\kappa}$ in (118),

$$\phi = \frac{\chi\delta}{\kappa\rho}. \tag{121}$$

This allows us to derive a variety of impulse responses. Calling y_t the process $dy_t = -\phi y_t dt + \sigma dz_t$, let us look for a solution of the form (or "Ansatz"):

$$p_t = A y_t, \qquad f_t = a y_t.$$

Plugging this Ansatz in (112) and examining the σz_t term gives:

$$a = 1.$$

Next, we have $D y_t = -\phi y_t$, so plugging this in (115) gives: $0 = \left[-\left(\zeta + \kappa\phi\right) A + a\right] y_t$, i.e.

$$A = \frac{1}{\zeta + \kappa\phi}.$$

This derived Proposition 9 in continuous time.□

One can derive other things. For instance, here's an expression for the price.

Proposition 11. (Equilibrium price in infinite horizon model, with enriched model of households)*Suppose that* $\Delta f_t = \chi \hat{\pi}_t + \Delta \tilde{f}_t$ *for some arbitrary* $\tilde{f}_t$*. Then, the price at time* t *is:*

$$p_t = \frac{f_{t-1}}{\bar{\zeta}} + \mathbb{E}_t \sum_{\tau=t}^{\infty} \frac{\rho}{\left(1+\rho\right)^{\tau-t+1}} \left(\frac{\tilde{f}_\tau - \tilde{f}_{t-1}}{\bar{\zeta}} + M^D d_\tau^e \right) \tag{122}$$

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 (119), and the price impact of an inflow $\Delta\tilde{f}_t$ decays at a rate ϕ.

12.3.2 Discrete time

We use the operator

$$\nabla x_t = \mathbb{E}_t \left[x_{t+1} - x_t\right]$$

so that $\nabla x_t \simeq D x_t \times \Delta t$. So $\hat{\pi}_t = -\delta p_t + \mathbb{E}_t \left[\Delta p_{t+1}\right]$ becomes:

$$\hat{\pi}_t = \left(\nabla - \delta\right) p_t.$$

Likewise, $\Delta f_t = \chi \hat{\pi}_t + \varepsilon_t$ (see (111)) implies:

$$\nabla f_t = \chi \hat{\pi}_{t+1} = \chi \left(\nabla - \delta\right) p_{t+1} = \chi \left(\nabla - \delta\right) \left(1 + \nabla\omega\right) p_t$$

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 (23) becomes:

$$0 = -\left(\zeta - \kappa\nabla\right) p_t + f_t.$$

Premultiplying by ∇ gives:

$$\begin{aligned} 0 &= -\nabla\left(\zeta - \kappa\nabla\right) p_t + \nabla f_t \\ &= -\nabla\left(\zeta - \kappa\nabla\right) p_t + \chi\left(\nabla - \delta\right)\left(1 + \nabla\omega\right) p_t \\ &= \tilde{H}\left(\nabla\right) p_t \end{aligned}$$

with

$$\tilde{H}(x) = (\kappa + \chi\omega) x^2 - (\zeta - \chi(1 - \delta\omega)) x - \chi\delta. \tag{123}$$

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:

$$p_t = Ay_t, \qquad f_t = ay_t. \tag{124}$$

This implies

$$\hat{\pi}_t = (\nabla - \delta) p_t = -(\phi + \delta) Ay_t.$$

Plugging this in (111) gives:

$$a = -\chi\omega(\phi + \delta) A + 1.$$

Plugging the Ansatz (124) in (23) gives:

$$0 = -(\zeta + \kappa\phi) A + a.$$

Hence, we obtain: $A = \frac{a}{\zeta + \kappa\phi}$, with

$$a = \frac{1}{1 + \chi\omega\frac{\phi+\delta}{\zeta+\kappa\phi}}. \tag{125}$$

Again, formally, we obtain the continuous time limit when $\omega \to 0$.

From discrete to continuous time Denote with bolded symbols the continuous-time version of the parameters, and Δt the calendar value of a time interval. Then, as they have units of $[\text{Time}]^{-1}$, we have

$$\phi = \boldsymbol{\phi}\Delta t, \qquad \delta = \boldsymbol{\delta}\Delta t,$$

but as it has unit of [Time], we have

$$\kappa = \boldsymbol{\kappa}/\Delta t \tag{126}$$

and there are unitless, χ and ζ are the same in discrete and continuous time.

Finally, we have $\omega = \Delta t$, as we wrote $\mathbb{E}_t[p_{t+1}] = (1 + \nabla\omega) p_t$. Hence, calling $x = \boldsymbol{x}\Delta t$,

$$\begin{aligned}\tilde{H}(x) &= (\kappa + \chi\omega) x^2 - (\zeta - \chi(1 - \delta\omega)) x - \chi\delta \\ \tilde{H}(x)/\Delta t &= (\boldsymbol{\kappa} + \chi\Delta t) \boldsymbol{x}^2 - (\zeta - \chi(1 - \delta\Delta t)) \boldsymbol{x} - \chi\delta\end{aligned}$$

so that indeed, $\lim_{\Delta t \to 0} \frac{\tilde{H}(x)}{\Delta t} = H(\boldsymbol{x})$.

12.3.3 Impact of a trend on dividends

We prove the following.

Proposition 12. *Suppose that $d_t = gt$. Then, if flows follow*

$$\Delta f_t = \chi\hat{\pi}_t + (1 - \theta) g + c$$

with $f_{-1} = 0$,*with* $\chi > 0$ *and some constant* c. *Then in the long run, the risk premium is higher,* $\hat{\pi}_* = -\frac{c}{\chi}$. *and we have* $p_t = d_t + p_*$, $f_t = (1-\theta) d_t + f_*$, *with* $p_* = \frac{c}{\chi\delta}$, *i.e.*

$$p_* = \frac{c}{\chi\delta} = \frac{c}{\kappa\phi\rho} \tag{127}$$

and $f = \zeta p_*$. *For finite* t, *we have*

$$p_t = d_t + \left(1 - \frac{\zeta}{\zeta + \kappa\phi} (1-\phi)^t\right) p_* \tag{128}$$

so that on impact

$$p_0 = \frac{c}{(\zeta + \kappa\phi)\rho} \tag{129}$$

where ρ,$-\phi$ *are the of the characteristic polynomial* $H(x)$ *in* (118). *The flows are*

$$f_t = (1-\theta) d_t + f_* \left(1 - (1-\phi)^t\right). \tag{130}$$

We write the rule as a deviation c from the rational flow, which is $\Delta f_t = (1-\theta) g$ by Lemma 1. In the baseline case $\Delta f_t = \chi\hat{\pi}_t + \varepsilon_t$, then $c = -(1-\theta) g < 0$. Intuitively, if there is a low χ, the "flows don't adjust enough", so that the price is too low, and the risk 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.[79]

Proof. First, we derive the long run, which is simpler. Calling $\hat{\pi}_*$ the steady state deviation of the risk premium from $\bar{\pi}$, we have on average $\Delta f_t = \chi\hat{\pi}_* + (1-\theta) g + c$. But Lemma 1 showed that we need $\Delta f_t = (1-\theta) g$. So, this implies $\hat{\pi}_* = \frac{-c}{\chi}$. This in turn corresponds to $p_t = p_* + gt$, with $p_* = -\frac{\hat{\pi}_*}{\delta}$.

Next, we derive the finite-time behavior. For simplicity, we use continuous time, and set $g = 0$ for simplicity (the general case is similar). We have $Df_t = \chi\hat{\pi}_t + c$. Insert this in (117) gives

$$H(D) p_t + c = 0. \tag{131}$$

The solution is $p_t = p_* + Ae^{-\phi t} + Be^{\rho t}$ for constants A and B. The large t behavior implies $B = 0$. As time $t = 0$, we must have $f_0 = 0$, so

$$0 = -\zeta p_0 + \kappa \frac{\mathbb{E} dp_0}{dt} = -\zeta p_* - (\zeta + \kappa\phi) A.$$

This gives $A = -\frac{\zeta}{\zeta + \kappa\phi} p_*$. This implies

$$p_0 = p_* + A = \left(1 - \frac{\zeta}{\zeta + \kappa\phi}\right) p_* = \frac{\kappa\phi}{\zeta + \kappa\phi} \frac{c}{\chi\delta}.$$

We use that $\phi = \frac{\chi\delta}{\kappa\rho}$ from (121), which gives (129).

[79] Note that it could be obtained in the case $\chi \to 0$ from Proposition 5, which gives $p_0 = \frac{c}{\zeta\rho}$ (by plugging in $f_\tau = c\tau$).

Finally, as $0 = -\zeta p_t + \kappa D p_t + f_t$, we have

$$f_t = \zeta p_* + (\zeta + \kappa\phi) A e^{-\phi t} = \zeta p_* \left(1 - e^{-\phi t}\right).$$

Using our calibration $g = 2\%$, $\theta = 0.85$ and rule (60) with $\chi = 0.23$ we find: $\hat{\pi}_* = \frac{1-\theta}{\chi} g = 1.2\%$, a moderate effect.

We next prove a result that synthesizes and expands on our previous results. □

Proposition 13. *Suppose an economy with i.i.d. dividend growth* $\Delta d_t = g + \varepsilon_t^d$*, and consumer flows following the semi-behavioral rule:*

$$\Delta f_t = \chi \hat{\pi}_t + (1 - \theta + \gamma) \Delta d_t + c + \varepsilon_t^f \tag{132}$$

where disturbances ε_t^d*,* ε_t^f *have mean 0 and no time correlations. The rational case obtains when* γ*,* χ*, c, var* $\left(\varepsilon_t^f\right)$ *are set to 0. Then, in the steady state, the risk premium is* $\bar{\pi} + \hat{\pi}$ *with* $\hat{\pi} = -\frac{\gamma g + c}{\chi}$*, and using* $p_* = -\frac{\hat{\pi}}{\delta}$*,* $f_* = \zeta p_*$*, and* ϕ *the mean-reversion of Proposition 9,*

$$f_t = (1 - \theta) d_t + \hat{f}_t + f_*, \qquad p_t = d_t + \frac{\hat{f}_t}{\zeta + \kappa\phi} + p_* \qquad \hat{f}_t = (1 - \phi) \hat{f}_{t-1} + \varepsilon_t^f + \gamma \varepsilon_t^d. \tag{133}$$

Here in (132), γ is a "gap", as the rational case would entail $\gamma = 0$. So, the "gap" creates a permanent change in the risk premium (as in Proposition 12). The new part is really the impact of disturbances ε_t^d, ε_t^f: their impact means-reverts at the rate ϕ. Excess flows ε_t^f make the price temporarily too high, and high dividends not immediately compensated by a flow ($\gamma \varepsilon_t^d$) make the price temporarily too low, as in the myopia effect of Proposition 3.

Proof. The terms corresponding to the non-zero trend g and c are exactly as in Proposition 12, using $c' = \gamma g + c$. So, by linearity, we can set $g = c = 0$ and focus on the stochastic terms. In the case where $\varepsilon_t^d = 0$, which is exactly Proposition 9. Then, the case ε_t^d is very similar, as the "mistake" in flows is the sum of the shock ε_t^f and the "excess adjustment" $\gamma \varepsilon_t^d$. □

12.4 Corporate finance in inelastic markets: Complements

We provide complements to Section 5.2.

12.4.1 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. (2007b) for an empirical analysis). So, the aggregate dividend goes from $\mathcal{D}_t$ to $\mathcal{D}'_t = \mathcal{D}_t (1 - b)$,and at each date corporations spend $b\mathcal{D}_t$ on share buybacks. To streamline the computations, we use continuous time.

Buybacks in a frictionless rational model We first consider the rational model.

Proposition 14. *Consider a firm that at time 0 changes its payout policy, and devotes a fraction b of the payout to share buybacks, $1-b$ to dividends (starting from paying only dividends before time 0). Consider a frictionless, rational model. Then, the dividend price ratio falls by a factor $1-b$:*

$$\delta' = (1-b)\,\delta.$$

It goes from $\delta = r+\pi-g$ to $\delta' = (1-b)\,\delta = r+\pi-g'$, where $g' = g+G$ if the new average growth rate of the dividend per share, which is increased by $G = b\delta$. At the same time, the market value of the firm is unchanged, as per Modigliani-Miller.

The surprise is that it changes the D/P ratio by a big amount. The share of dividend as a fraction of the payout has moved from roughly 100% to 50%, so $b = 0.5$. Hence, Proposition 14 implies that the price-dividend ratio went from δ (empirically, about 4%) to half its value (about 2%). This is simply because the growth rate per share has increased by $G = 2\%$, because the number of shares has decreased by the same $G = 2\%$.

Proof. We had $P_t Q_t = \frac{\mathcal{D}_t}{\delta}$. As $b\mathcal{D}_t$ dollars are devoted to purchasing shares each period, the number of shares follows $\dot{Q}_t = -\frac{b\mathcal{D}_t}{P_t} = -bQ_t\delta$, so that the new number of shares is:

$$Q'_t = Q_0 e^{-Gt}, \qquad G = b\delta. \tag{134}$$

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

$$D'_t = (1-b)\,e^{Gt} D_t. \tag{135}$$

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

$$P'_t = P_t e^{Gt}. \tag{136}$$

This implies that the new D/P ratio is

$$\frac{D'_t}{P'_t} = \frac{D_t}{P_t}(1-b)\,. \tag{137}$$

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 (134) in the last equation. □

Buybacks in an inelastic model We next study the situation in an inelastic model.

Proposition 15. (Impact of share buybacks in the infinite horizon model) *In the inelastic model, suppose that a change in policy on a scale b is announced, and should last forever. Then, with $c = \left(\mu^D - \mu^G\right)\theta\delta b$, the firm value increases on impact by $v_0 = \frac{c}{(\zeta+\kappa\phi)\rho}$ and in the long run by $v_* = \frac{c}{\kappa\phi\rho}$. In between, the change in firm value is $v_t = v_* + (v_0 - v_*)\,(1-\phi)^t$.*

Hence, the economics is similar to the simple model of Proposition 8.

Proof. As in the rational case, growth rate in the number of shares is $-G$ with $G = \delta b$, $d_t = -b + Gt$. As at each period the dividends are lower by b, and they represented a fraction $\delta\theta$ of the fund's value, $f^{C\to M} = -\int_0^t \delta b\theta dt = -\theta Gt$, and $q^S = -Gt$. So, if households do not change their flows ($\mu^D = \mu^G = 0$), by Proposition 4, we have $p_t = Gt$. Hence, the value of the firm is unchanged.

Now, consider the case where households do change their flows: as each period the dividends are lower by b, and they represented a fraction $\delta\theta$ of the fund's value, and capital gains are increased by G, we have

$$\Delta f^{h\to M} = -\mu^D(-b\delta\theta) - \mu^G G\theta = \left(\mu^D - \mu^G\right)\theta G.$$

Equilibrium is

$$-\zeta p_t + \kappa \mathbb{E}\Delta p_{t+1} + f^{C\to M} + f^{h\to M} = q^S$$

i.e.

$$-\zeta p_t + \kappa \mathbb{E}\Delta p_{t+1} + f_t = 0$$

with

$$f = -q^S + f^{C\to M} + f^{h\to M} = \left(\mu^D - \mu^G\right)\theta Gt + (1-\theta)\,Gt.$$

So, we are in the situation of Proposition 12, with $g = G = \delta b$, $c = \left(\mu^D - \mu^G\right)\theta G$. Hence, the value of the firm increases on impact by:$p_0 = v_0 = \frac{c}{(\zeta+\kappa\phi)\rho}$ and in the long run by $v_* = p_* = \frac{c}{\kappa\phi\rho}$. □

Tentative calibration We proceed as above, with $\mu^D = 0.5$, $\mu^G = 0.03$. Under this tentative calibration, a buyback of 1% of the market increase the market capitalization by 0.6% in the short run, and 1.6% in the long run.

12.4.2 An increase in buyback financed by increased debt rather than lower contemporaneous dividends.

Consider an increase in buybacks that is not compensated by a contemporaneous decrease in dividends — so that the total payout is increased, by a factor equal to b times the initial market value.

Two period model. The buyback of B dollars decreases the number of shares by $\frac{B}{P_0}$, the future aggregate dividend by BR, where R is the gross interest rate. The time-1 dividend is $\mathcal{D}_1' = \mathcal{D}_1 - BR$, the present value of $\mathcal{D}_1'$ falls by a fraction b. As the number of shares also falls by b, the present value of the time-1 dividend per share remains constant:

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

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

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

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

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

Hence, the aggregate values of equities increases, and the time-0 return r is positive, unless it's compensated by a flow $f^h = -b$. Using the marginal propensities in Section 5.2, we have $f^h = -\mu^G \theta p$, so that in total:

$$p = \frac{b}{\zeta + \theta \mu^G}. \tag{138}$$

Infinite horizon. We suppose that the debt will be repaid very far in the future (at date $T \to \infty$.) Then, the economics is as in the two-period model.[80]

12.5 General equilibrium in a two-period economy

We detail a two-period endowment economy. This clarifies a number of things, such as the link between consumption and cash-flows, and the determination of the interest rate.

There are a representative consumer and a representative firm in an endowment economy. The consumer has utility

$$U(C_0, C_1) = u(C_0) + \beta u(C_1). \tag{139}$$

In equilibrium all consumption at time $t \in \{0, 1\}$ comes from a non-storable endowment Y_t. The time-0 endowment Y_0 has been given to the consumer as a time-0 dividend. Output Y_1 is stochastic, with a strictly positive lower bound.

The fruit from the tree goes to the representative firm, which pays households as debt repayment and stock dividend, Y_1.

At time 0, the government issues a bond in quantity Q^B, and gives it to the citizens. It is backed by taxes $R_f Q^B$ collected at time 1. This does not affect aggregate consumption, but it clarifies the origins of bonds. One could also just set $Q^B = 0$. Then, the pure bond fund will hold a negative quantity of bonds.

We call $R_1^{\mathcal{E}} = \frac{Y_1}{P_0}$ the gross return on equities. If the consumer sells I of bonds to invest them in equities, and saves an additional s dollars (so that $C_0 = Y_0 - s$), time-1 wealth is:

$$W_1 = \left(P_0^{\mathcal{E}} Q^{\mathcal{E}} + I\right) R_1^{\mathcal{E}} + \left(Q^B - I + s\right) R_f - Q^B R_f.$$

The first term is the proceeds from equities, the second term is the proceeds from the pure bond fund, and the last term the taxes paid to the government. This simplifies to:

$$W_1(I, s) = \left(P_0^{\mathcal{E}} Q^{\mathcal{E}} + I\right) R_1^{\mathcal{E}} + (-I + s) R_f. \tag{140}$$

A traditional, rational benchmark Let us examine what would happen in the traditional, rational benchmark. The consumer's problem would be

$$\max_{s,I} u(Y_0 - s) + \mathbb{E}\left[u(W_1(I, s))\right].$$

In equilibrium, we should have $s = I = 0$. Hence, Lucas (1978) pricing would hold. We would have a rational pricing kernel:

$$\mathcal{M} = \beta \frac{u'(Y_1)}{u'(Y_0)}. \tag{141}$$

[80] In a rational model, we still have $p_t = 0$, $v_t = -b$. In an inelastic model, as $\rho > \delta$,we have $d_t = 0$, $q_t^S = -b$, so $p_t = \frac{b + f^h}{\zeta}$ (for all dates $t \ll T$) also. Hence the expression is as in the two-period model.

The interest rate R_f would satisfy

$$R_f \mathbb{E}[\mathcal{M}] = 1, \tag{142}$$

and the time-0 price of equities would be:

$$P_0^* = \mathbb{E}[\mathcal{M} D_1]. \tag{143}$$

But we wish to propose an alternative to that Lucas (1978) approach.

As in the simple model of Section 3, in addition to a pure bond vehicle there is a balanced allocation fund that invests in a mix of bonds and equities. The market value of the mixed fund is $W_t^M = Q_{Mt}^B + Q_{Mt}^{\mathcal{E}} P_t$ at dates $t = 0^-$ and $t = 0$. Its value at time $t = 1$ is $W_1^M = Q_{M0}^B R_f + Q_{M0}^{\mathcal{E}} D_1$.

We assume that we are in equilibrium at time 0^-, before the flow and dividend news. By this we mean that households have allocated their holdings to the bond and mixed fund in the right proportion, so that equity values are correct, $P_{0^-} = P_0^*$ as in (143). As the mixed fund allocates a fraction θ in equity, and $1-\theta$ in bonds, it manages a wealth $W_{M,0^-} = \frac{1}{\theta} Q^{\mathcal{E}} P_0^*$, and a quantity of bonds $Q_{M,0^-}^B = \frac{1-\theta}{\theta} Q^{\mathcal{E}} P_0^*$. The pure bond fund holds the remaining bonds, $Q^B - Q_{M,0^-}^B$.

The household is endowed with all shares of the pure bond fund and mixed fund. He may wish to increase his savings by s; and he may wish to invest an additional ΔF dollars in the mixed fund, taken from selling ΔF from the pure bond fund. So, the consumptions are:

$$C_0 = Y_0 - s, \qquad C_1 = Y_1 + sR_f + \left(\frac{W_1^M}{W_0^M} - R_f \right) \Delta F,$$

which defines $C_t(s, \Delta F)$ and the expected utility

$$U(s, \Delta F) = u(C_0(s, \Delta F)) + \beta \mathbb{E}[u(C_1(s, \Delta F))].$$

If the household were rational, its plan would be to maximize U subject to the above budget constraint: $\max_{s,\Delta F} U(s, \Delta F)$.

Flows and behavioral assumptions We model households as rational in consumption, but behavioral in portfolio choice. More formally, we decompose the household as a consumer, who can only trade the pure bond fund (so, chooses s) and decide on consumption (given the money in the pure bond fund), and an investor, who trades the different funds (so chooses ΔF), but cannot decide on consumption.

The consumer is rational, and she "sees through" all the cash-flows. Her plan is $\max_s U(s, \Delta F)$. She sees that in equilibrium she will consume the endowment Y_t at time t. As she trades the bond, she enforces the Euler equation, so that (142) holds with pricing kernel (141).

Suppose that at time 0 there is a shock to the expectation of the future output Y_1 and a flow shock ΔF. To concentrate on the essentials, we assume that this shock does not change the frictionless interest rate R_f. Then, the economy is as described in Section 3. In particular, the price changes as in that simple model.

12.6 Details of the household's problem in general equilibrium

This section provides some extra details to the household's problem of Section 6.1.

We call t^- the beginning of period values, evaluating all at the time t price P_t. The mixed fund gives a dividend $D_t^M = Q^{\mathcal{E}} Y_t + r_{f,t-1} B_{t-1}^M$, so that its cum-dividend value is $W_{t^-}^M = Q^{\mathcal{E}} P_t + B_{t-1}^M + D_t^M$, and the return is $R_t^M = \frac{W_{t^-}^M}{W_{t-1}^M}$.

The mixed fund has issued N_{t-1} shares, of which N_{t-1}^h are owned by household h. The value of a share in the mixed fund is $v_t^M = \frac{Q^{\mathcal{E}} P_t + B_{t-1}^M}{N_{t-1}}$. So, the beginning of period wealth of the household is:

$$W_{t^-}^h = \frac{N_{t-1}^h}{N_{t-1}} W_{t^-}^M + B_{t-1}^h R_{f,t-1}. \tag{144}$$

Suppose the households flow ΔF_t^h into the mixed fund, while the rest of the economy flows ΔF_t (in equilibrium, the two values are the same). Then, the number of shares owned by the household and in the fund are:[81] $N_t^h = N_{t-1}^h + \frac{\Delta F_t^h}{v_t^M}$ 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:

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

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

12.7 More general cases to get a pricing kernel

Here we expand Section 5.3, to multiple risky assets and a consumption-based default SDF.

A Gaussian example To be clear let us work out a basic example. We suppose that returns and consumption are Gaussian:

$$\frac{C_1}{C_0} = e^{g_c + \sigma_c \varepsilon^c - \frac{1}{2}\sigma_c^2}, \tag{145}$$

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

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

for the risk-free rate $r_f = -\ln \beta + \gamma g_c - \frac{1}{2}\gamma\left(1+\gamma\right)\sigma_c^2$.

We next consider the agile optimizers' problem, going back to a general default pricing kernel $\mathcal{M}^d$ (that might be $\mathcal{M}^{d,R_f}$ or $\mathcal{M}^{d,C}$). We recall that for two jointly Gaussian variables X, Y:

$$\frac{\mathbb{E}\left[e^X Y\right]}{\mathbb{E}\left[e^X\right]} = \mathbb{E}\left[Y\right] + cov\left(X, Y\right). \tag{146}$$

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

For instance, the anomalous excess risk premium is

$$\mathbb{E}^{\mathcal{M}^d}[R] := \frac{\mathbb{E}\left[\mathcal{M}^d R\right]}{\mathbb{E}\left[\mathcal{M}^d\right]} = \mathbb{E}[R] + cov\left(\mathcal{M}^d, R\right), \tag{147}$$

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

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

where $-cov\left(\mathcal{M}^d, R\right)$ is the risk premium warranted by the default pricing kernel. The optimal portfolio of agile optimizers is $Q = V_R^{-1}\mathbb{E}^{\mathcal{M}^d}[R]$ and their return is a form of "tangency portfolio" return:

$$R^\tau = Q'R = \mathbb{E}^{\mathcal{M}^d}[R]' V_R^{-1} R, \tag{148}$$

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

$$\mathcal{S} = \frac{\mathbb{E}^{\mathcal{M}^d}[R^\tau]}{\sigma_{R^\tau}}, \tag{149}$$

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

$$\mathcal{M} = \mathcal{M}^d e^{-\mathcal{S}\frac{R^\tau - \mathbb{E}^{\mathcal{M}^d}[R^\tau]}{\sigma_{R^\tau}} - \frac{1}{2}\mathcal{S}^2}. \tag{150}$$

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

12.8 On the link between the Kyle lambda and the market inelasticity

Suppose that within a time window, there is an "order flow" (realized signed trades), with volume x_t, expressed as a fraction of the market capitalization. A typical microstructure regression is:

$$p_t - p_{t-1} = \lambda\left(x_t - \mathbb{E}_{t-1}[x_t]\right) \tag{151}$$

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 risk premium on average, whereas in our model, positive inflows lower the risk premium.

To analyze what happens in our model, we suppose some autocorrelation in the order flow (like Madhavan et al. (1997) and Bouchaud et al. (2018)):

$$x_t = (1 - \phi_x) x_{t-1} + \varepsilon_t,$$

where ε_t is i.i.d. So, an innovation ε_t creates an innovation to the cumulative expected flow $X_T = \sum_{\tau=0}^{T} x_\tau$ (for some very large T), equal to[82]

$$X_t = \mathbb{E}\left[X_T | I_{t-1}, x_t\right] - \mathbb{E}\left[X_T | I_{t-1}\right] = K\varepsilon_t,$$

where $K = \frac{1}{\phi_x}$. For instance, if a large desired trade ("meta-order") is "sliced" into 10 trades on average, then $K = 10$. Likewise, if a fast institution trades, and is followed on average by 4 more similar trades by other, slower institutions, then $K = 5$.[83]

In our model, the total price impact is $p_t - p_{t-1} = \frac{1}{\zeta} X_t$, which is

$$p_t - p_{t-1} = \frac{1}{\zeta} X_t = \frac{K}{\zeta} \varepsilon_t = \frac{K}{\zeta} \left(x_t - \mathbb{E}_{t-1}\left[x_t\right]\right).$$

Hence, an econometrician estimating (151), will find:

$$\lambda = \frac{K}{\zeta}. \tag{152}$$

This means that, for the aggregate market, the Kyle lambda is the inelasticity $\frac{1}{\zeta}$ times the persistence parameter K associated with the positive autocorrelation of the order flow.

Most empirical work in microstructure is done at the level of one asset, so that the λ they estimate is

$$\lambda = \frac{K}{\zeta^{\perp}}, \tag{153}$$

where $\lambda^{\perp}$ is the micro-elasticity of Section 3.6. In practice, one indeed finds that λ (e.g. as estimated byFrazzini et al. (2018)) is much greater than $\frac{1}{\zeta^{\perp}}$, with a coefficient $K \simeq 10$.

12.9 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 will find that they are quantitatively small (which makes sense, as prices are hard to forecast), so that we recommend that the reader skip this subsection at first.

12.9.1 Modeling a fund's inertial adjustment

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:

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

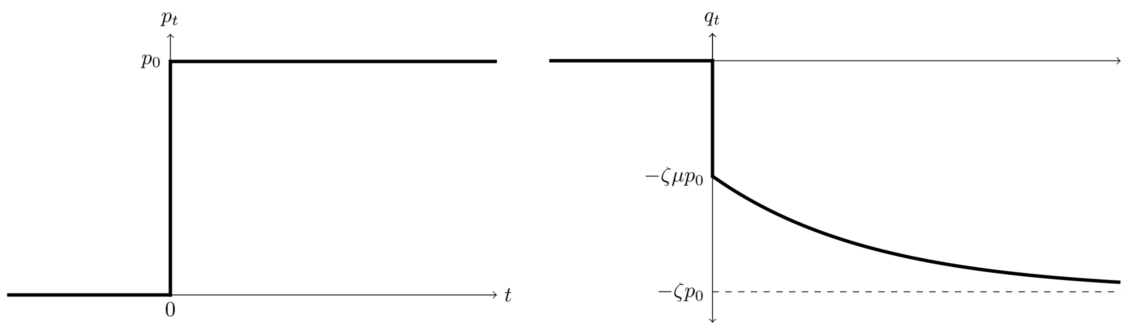
We model the actual active demand with inertia as:

$$\Delta q_t^a = \mu \Delta q_t^{a,\nu} + \phi \Delta t \left(q_{t-1}^{a,v} - q_{t-1}\right), \tag{154}$$

[82]Really, by Proposition 5, future flows should be discounted. But here we take the limit of fairly fast trading, so $\phi \gg \rho$, so that it is as if they were not discounted. This lightens up the notation, without changing the core economics.

[83]More generally (e.g. in models with multiple time scales), K is the "expected value of related orders, given the past", rather than the inverse mean-reversion.

Figure 14: This figure shows the quantity adjustment of an inert fund after a change in the aggregate stock price. It illustrates Proposition 16. If investors are inertial, there is a gradual adjustment of the quantity over time. When there is no inertia, $\mu = 1$ and the quantity adjustment is instantaneous.



(a) Impulse

(b) Fund's Quantity Response

where $q_t^{a,v}$ is the "virtual active demand" – the one of a non-inertial fund:

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

with $\mu \in [0,1]$ and $\phi \geq 0$. A very frictional investor has $\mu = 0$ and $\phi \geq 0$ low. A frictionless investor has $\mu = 1$. More frictional investors have lower $\mu \geq 0$ and lower $\phi \geq 0$. The adjustment to flows f_t is instantaneous for simplicity, and as it does not require a "strategic" decision by the fund.

It is useful to write those in continuous time:

$$q_t^{a,v} = -\zeta p_t + \kappa \mathbb{E}_t \dot{p}_t + \kappa\delta d_t - \bar{\pi} + \nu_t, \quad (155)$$

$$\dot{q}_t^{a} = \mu \dot{q}_t^{a,\nu} + \phi\left(q_t^{a,v} - q_t\right). \quad (156)$$

12.9.2 Quantity dynamics when funds are inertial

We derive quantity adjustments.

Proposition 16. (Short-run versus long-run elasticity of demand) *Suppose a fund that exhibit inertia. Then, its short-run elasticity of demand is $\mu\zeta$, and its long-run elasticity of demand is ζ. More precisely, suppose that the log price of equities jumps by p_0 at time 0, i.e. $p_t = 1_{t\geq 0} p_0$. Then, at $t \geq 0$, the fund's demand change is:*

$$q_t = -\zeta\left(1 + (\mu - 1)\, e^{-\phi t}\right) p_0, \quad (157)$$

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

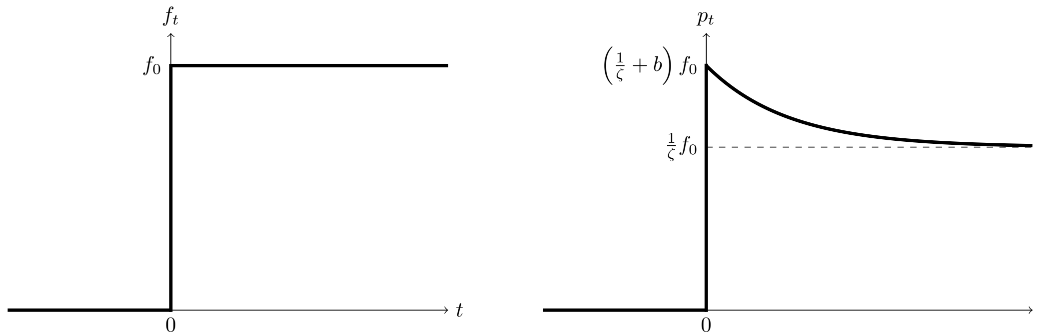
Figure 14 illustrates the dynamics.

12.9.3 Price dynamics when funds are inertial: short-run versus long-run elasticities

Suppose that there's an inflow f_t. What's the price impact?[84]

[84] Section 12.9.4 generalizes the price as present value formula (20) and the price impact with inertia formula (158) to the case of flows happening at any date.

Figure 15: This figure shows the price dynamics caused by an unanticipated time-0 demand shock, when investors are inertial. It illustrates Proposition 17. The permanent demand shock f_t creates a permanent price change $p_\infty = \frac{f_0}{\zeta}$. If investors are inertial, there is a small extra bump b on impact, that decays exponentially over time. When investor are not inert, $b = 0$ and the price immediately jumps to its permanent value p_∞.



(a) Impulse

(b) Price Response

Proposition 17. (Price impact with inertial inflow) *When funds exhibit inertia, the price impact of a permanent, unanticipated inflow f_0 at time 0 is (for $t \geq 0$),*

$$p_t = \left(\frac{1}{\zeta} + be^{-\Phi t}\right) f_0 \tag{158}$$

where

$$\Phi = \frac{\phi}{\mu}, \tag{159}$$

$$b = \frac{1-\mu}{\mu\left(\zeta + \kappa\Phi\right)}. \tag{160}$$

Proof. We conjecture a solution of the type (158). We normalize $f_0 = -1$. Plugging this in (155) gives

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

We should have $q_t^a + f_0 = 0$, i.e. $q_t^a = 1$ for $t > 0$. So (156) gives:

$$0 = \dot{q}_t^a = \mu \dot{q}_t^{a,\nu} + \phi\left(q_t^{a,v} - q_t\right) = c\left(-\Phi\mu + \phi\right) e^{-\Phi t},$$

which leads to $\Phi = \frac{\phi}{\mu}$. At time 0, (156) gives

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

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

Figure 15 illustrates the dynamics of (158). An unanticipated, permanent inflow f_0 at time 0 has an immediate price impact $\left(\frac{1}{\zeta} + b\right) f_0$ that is slightly bigger than the long-run price impact $\frac{f_0}{\zeta}$. The initial "excess reaction" bf_0 dies down at the exponential rate Φ. When funds are not inert, $b = 0$.

For $\mu < 1$, (159) implies that $\Phi > \phi$: the speed of price dynamics Φ is faster than the fund-level speed of adjustment of quantities ϕ. The intuition is as follows. Imagine that the impulse is a positive inflow, which increases the price. First, the "active" part of the fund strategy wants to sell shares, as the price is high and the risk 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.

Heterogeneity in inertia across funds. One can generalize this 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 (158) holds, with more complex expressions for b and Φ.

Impact of inertia on price predictability: calibration 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 (159) is $\Phi = \frac{\phi}{\mu} = 2/\text{year}$, and the overshooting of flows on impact in (160) is, using $\kappa = 1$ for illustration:

$$b = \frac{1-\mu}{\mu\zeta + \phi\kappa} = \frac{1-0.5}{0.5 \cdot 0.2 + 1 \cdot 1.5} \simeq 0.3.$$

The immediate price impact is $\frac{1}{\zeta} + b = 5.3$, while the permanent price impact is $\frac{1}{\zeta} = 5$. So, the temporary bump $b \ll \frac{1}{\zeta}$ is pretty negligible in the big picture. As the price decays as $bf_t e^{-\Phi t}$, the premium is $b\Phi f = 0.3 \cdot 2 \cdot 0.5\% = 0.3\%$ (if $f_t = 0.5\%$), again a small premium.

12.9.4 Impact of anticipated and unanticipated flows when funds are partially inert

The next proposition generalizes the price as present value formula (20) and the price impact with inertia (158).

Proposition 18. (Price impact with inertial funds) *When funds exhibit inertia, the price impact of inflows df_s is:*

$$p_t = \frac{f_{-\infty}}{\zeta} + \int_{-\infty}^{\infty} G(t-s)\, \mathbb{E}_t [df_s], \tag{161}$$

where

$$G(\tau) = \begin{cases} \left(\frac{1}{\zeta} + b\right) e^{\rho\tau} & \text{if } \tau \le 0, \\ \frac{1}{\zeta} + be^{-\Phi\tau} & \text{if } \tau \ge 0. \end{cases} \tag{162}$$

When there is no inertia, $b = 0$.

Proof. This can be checked by the "plug and verify" method. □

References for Online Appendix

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.

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.

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

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

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.

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.