NBER WORKING PAPER SERIES

IN SEARCH OF THE ORIGINS OF FINANCIAL FLUCTUATIONS: THE INELASTIC MARKETS HYPOTHESIS

Xavier Gabaix
Ralph S. J. Koijen

Working Paper 28967
http://www.nber.org/papers/w28967

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2021

We thank Ehsan Azarmsa, Aditya Chaudhry, Antonio Coppola, Zhiyu Fu, Dong Ryeol Lee, Hae-Kang Lee, Simon Oh, and Lingxuan Wu for excellent research assistance. We thank Francesca Bastianello, Jean-Philippe Bouchaud, Michael Brandt, John Campbell, Francesco Franzoni, Robin Greenwood, Valentin Haddad, Lars Hansen, Sam Hanson, John Heaton, Tim Johnson, Arvind Krishnamurthy, Spencer Kwon, John Leahy, Hanno Lustig, Alan Moreira, Knut Mork, Toby Moskowitz, Stefan Nagel, Jonathan Parker, Lasse Pedersen, Joel Peress, Jean-Charles Rochet, Ivan Shaliastovich, Andrei Shleifer, Jeremy Stein, Johannes Stroebel, Larry Summers, Adi Sunderam, Jean Tirole, Harald Uhlig, Dimitri Vayanos, Motohiro Yogo, and participants at various seminars for comments. Gabaix thanks the Sloan Foundation for financial support. Koijen acknowledges financial support from the Center for Research in Security Prices at the University of Chicago Booth School of Business. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis
Xavier Gabaix and Ralph S. J. Koijen
NBER Working Paper No. 28967
June 2021
JEL No. E7,G1,G32,G4

ABSTRACT

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

Using the recent method of granular instrumental variables, we find that investing $1 in the stock market increases the market's aggregate value by about $5. We also develop a new measure of capital flows into the market, consistent with our theory. We relate it to prices, macroeconomic variables, and survey expectations of returns.

We analyze how key parts of macro-finance change if markets are inelastic. We show how general equilibrium models and pricing kernels can be generalized to incorporate flows, which makes them amenable to use in more realistic macroeconomic models and to policy analysis.

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

Xavier Gabaix
Department of Economics
Harvard University
Littauer Center
1805 Cambridge St.
Cambridge, MA 02138
and NBER
xgabaix@fas.harvard.edu

Ralph S. J. Koijen
University of Chicago
Booth School of Business
5807 S Woodlawn Ave
Chicago, IL 60637
and NBER
Ralph.koijen@chicagobooth.edu

An online appendix is available at http://www.nber.org/data-appendix/w28967

1 Introduction

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

We start by asking a simple question: when an investor sells \$1 worth of bonds and buys \$1 worth of stocks, what happens to the valuation of the aggregate stock market? In the simplest "efficient markets" model, the price is the present value of future dividends, so the valuation of the aggregate market should not change. However, we find both theoretically and empirically, using an instrumental variables strategy, that the market's aggregate value goes up by about \$5 (our estimates are between \$3 and \$8, and we will use \$5 for simplicity in the theory and discussion).[1] Hence, the stock market in this simple model is a very reactive economic machine, which turns an additional \$1 of investment into an increase of \$5 in aggregate market valuations.

Put another way, if investors create a flow of 1% as a fraction of the value of equities, the model implies that the value of the equity market goes up by 5%. This is the mirror image of the low aggregate price-elasticity of demand for stocks: if the price of the equity market portfolio goes up by 5%, demand falls by only 1%, so that the price elasticity is 0.2. In contrast, most rational or behavioral models would predict a very small impact, about 100 times smaller, and a price elasticity about 100 times larger. This high sensitivity of prices to flows has large consequences: flows in the market and demand shocks affect prices and expected returns in a quantitatively important way. We refer to this notion as the *"inelastic markets hypothesis."*

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

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

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

[1] The price impact is linear and symmetric: selling \$2 worth of equities (buying \$2 worth of bonds) decreases the valuation of aggregate equities by \$10.

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

Importantly, the data are consistent with a quite long-lasting price impact of flows. Indeed, in the simplest version of the model, the price impact is perfectly long-lasting. This is not necessarily because flows release information, but instead simply because the permanent shift in the demand for stocks must create a permanent shift in their equilibrium price. We perform a large number of robustness checks, for example using different data sets (the Flow of Funds as well as 13F filings). The findings are consistent across specifications, in the sense that the price impact multiplier remains around 5. We also construct a measure of capital flows into the market. We find that this measure is strongly correlated with realized returns and survey expectations of returns, but it is only weakly correlated with macroeconomic growth.

Here are three a priori reasons to entertain that markets would be inelastic First of all, if one wants to buy $1 worth of equities, many funds actually cannot supply that: for instance, a fund that invests entirely in equities cannot exchange them for bonds. Many institutions have tight mandates, something that we confirm empirically. Relatedly, it is hard to find investors who could act as macro arbitrageurs. For instance, hedge funds are relatively small (they hold less than 5% of the equity market), and they tend to reduce their equity allocations in bad times (due to outflows and binding risk constraints; see Ben-David et al. (2012)). Second, the transfer of equity risk across investor sectors is small (about 0.6% of the aggregate value of the equity market per quarter for the average pair of investor sectors). This implies that the demand elasticity of most investors is quite small or that investors experience nearly identical demand shocks (as if they were to disagree, we would see large flows in elastic markets), something which may be implausible. Third, a large literature estimates demand elasticities for individual stocks using a variety of methodologies, where the latest estimates of this "micro" demand elasticity are approximately 1 (we provide complete references below). As the macro elasticity should arguably be lower than the micro elasticity (considering that, for example, Ford and General Motors are closer substitutes than the stock market index and a bond), this suggests a low macro elasticity, perhaps less than 1. Consistent with this reasoning, a new literature explores elasticities for "factors" in the US, such as size and value, and finds elasticities of around 0.2. Hence, in light of this existing evidence, our low macro elasticity may be less surprising.

Suppose that the "inelastic markets hypothesis" is true; why do we care? First, investor-specific flows and demand shocks are quantitatively impactful. As a result, one can replace the "dark matter" of asset pricing (whereby price movements are explained by hard-to-measure latent forces) with tangible flows and the demand shocks of different investors. This suggests a research program in which determinants of asset prices can be traced back to measurable demand shocks and flows of concrete investors. By studying the actions of these investors, we can infer their demand curves, and theorize about their determinants.

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

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

If markets are inelastic, then macro-finance should reflect that. Accordingly, we construct a general equilibrium model in the spirit of Lucas (1978) where there is a central role for flows and inelasticity. It clarifies the role of demand shocks and flows, the determination of the interest rate, and shows how to augment traditional general equilibrium models with flows in inelastic markets. That makes those models more realistic, and better suited for policy. This model may serve as a prototype for models enriched by inelasticity. Indeed, it calibrates well, and replicates quantitatively the salient features of the stock market, such as the volatility and size of the equity premium, the slow mean-reversion of the price-dividend ratio, and the ability to predict stock return with the price-dividend ratio at different horizons. We also show how the model can be used to match the strong correlation between prices and subjective beliefs about long-term growth (Bordalo et al. (2020)), even if fluctuations in beliefs have only a modest impact on actions (Giglio et al. (2021a)), as the resulting flows are amplified in inelastic markets. We conclude that our general equilibrium model with "inelastic markets" is competitive with other widely-used general equilibrium models that match equity market moments, be it via habit formation (Campbell and Cochrane (1999)), long run risks (Bansal and Yaron (2004)), or variable rare disasters (Gabaix (2012), Wachter (2013)). In addition to proposing a new amplification mechanism, its main advantage, as we see it, is that is relies on an observable force, flows in and out of equities.

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

One limitation of our study is that we postpone to future research the detailed investigation of what determines flows in the first place: instead, we provide descriptive statistics showing they

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

correlate sensibly with other variables, such as prices and measured beliefs. The reason is chiefly that this would be a stand-alone paper. But we think it is quite doable, and indeed we are working on this. Rather than studying "shocks to noise traders" abstractly, we replace them with investor-level flows and demand shocks that may be easier to understand. Indeed, episode by episode, one can ask questions such as "why did firms lower their buybacks?" (answer: because they had lower earnings), "why did pension funds buy?" (answer: because their mandate forces them to buy stocks after stocks fall), or "why did hedge funds sell?" (answer: their investors sold, given their low past returns).

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

We build on the insights of De Long et al. (1990), who write an equilibrium model in which noisy beliefs create demand shocks that move the market and the equity premium. They discuss a rich set of qualitative ideas, some of which we can formally analyze and quantify, such as the failure of the Modigliani-Miller theorem and the notion that if most market participants passively hold the market portfolio, prices react sharply to flows. De Long et al. (1990) dealt with these issues qualitatively, but, influenced by it, a literature has studied the impact of mutual fund flows in the market, for example Warther (1995).[4] In addition, an active literature studies the impact of mutual fund and ETF flows on the cross-section of equity prices, for instance Frazzini and Lamont (2008), Lou (2012), Ben-David et al. (2018), Dou et al. (2020), and Dong et al. (2021). One innovation of our paper is to provide a systematic quantitative framework to think about this, to include all sectors (not just mutual funds), and to think about causal inference at the level of the aggregate stock market via GIV. Deuskar and Johnson (2011a) use high-frequency order flow data for S&P 500 futures to show that about half of the price variation can be attributed to flows shocks. Moreover, they find these shocks to be permanent over the horizons that they consider.[5]

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

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

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

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

Gabaix and Koijen (2020).

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

Our work also relates back to the work on flows and asset demand systems by Brainard and Tobin (1968) and Friedman (1977), among others. This literature faced two important challenges that we address; first, data on asset holdings were not as readily available as they are now and, second, there were no obvious methods to identify the slopes of asset demand curves. We share with Koijen and Yogo (2019) and Koijen et al. (2019) our reliance on holdings data by institutions, and the desire to estimate a demand function. We are mostly interested in the equilibrium in the aggregate stock market, as opposed to the cross-sectional focus of Koijen and Yogo (2019), and we emphasize the role of flows, and the dynamics of prices and capital flows over time. Using a similar modeling strategy as in Koijen and Yogo (2019), Koijen and Yogo (2020) estimate a global demand system across global equity and bond markets to understand exchange rates, bond prices, and equity prices across countries. We also relate to the literature on slow-moving capital (Mitchell et al. (2007); Duffie (2010); Duffie and Strulovici (2012); Moreira (2019); Li (2018)), providing a new model for price impact with long-lasting effects, and an identified estimation. Finally, part of our contribution is a new model of intermediaries (He and Krishnamurthy (2013)), with a central role for flows, trading mandates, and inelasticity.

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

Outline Section 2 gives some simple suggestive facts on equity shares and potential macro arbitrageurs such as broker dealers and hedge funds. It also summarizes the existing literature on elasticity estimates. Section 3 develops our basic model of the stock market: it lays out the basic notions, and defines clearly elasticity and its link with price impact. It also gives the theoretical framework that we take to the data. Section 4 contains the empirical analysis, including with an instrumental variable estimation of the aggregate market elasticity. Section 5 provides a general equilibrium model that helps to think about how everything fits together: it specializes the basic model of Section 3 as it endogenizes the interest rate and links cash flows to production and consumption. Section 6 discusses how the effectiveness of government policy and corporate finance change with inelastic markets. Section 7 provides a conclusion and thoughts about the research directions suggested by the present approach. The appendix contains the basic proofs, and details. The online appendix contains a number of robustness checks and extensions.

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

$\sum_{i=1}^{N} S_i = 1$, we let $X_E := \frac{1}{N}\sum_{i=1}^{N} X_i$, $X_S := \sum_{i=1}^{N} S_i X_i$, $X_\Gamma := X_S - X_E$ so that X_E is the equal-weighted average of the vector's elements, X_S is the size-weighted average, and X_Γ is their difference. We define the mean of X_i (with $i = 1 \dots N$) with weights ω_i as: $\mathbb{E}_\omega [X_i] := \frac{\sum_i \omega_i X_i}{\sum_i \omega_i}$.

2 Data and Suggestive Facts on Equity Shares and Flows

In this section we document several stylized facts and discuss how they are related to our model and to traditional, elastic asset pricing models. These facts are meant to be no more than suggestive: the core empirical results are in Section 4, in which we try to carefully quantify the key parameters of our model.

After discussing the data construction in Section 2.1, we document that institutions often have quite stable equity shares in Section 2.2, and relatedly we seek to identify investors with elastic demand for the aggregate stock market in Section 2.3. That is, we ask: who are the deep-pocketed arbitrageurs that could make the aggregate stock market elastic? This question relates to the work by Brunnermeier and Nagel (2004), who show that hedge funds did not provide elasticity to the market during the technology bubble in the late nineties.

2.1 Data sources and construction

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

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

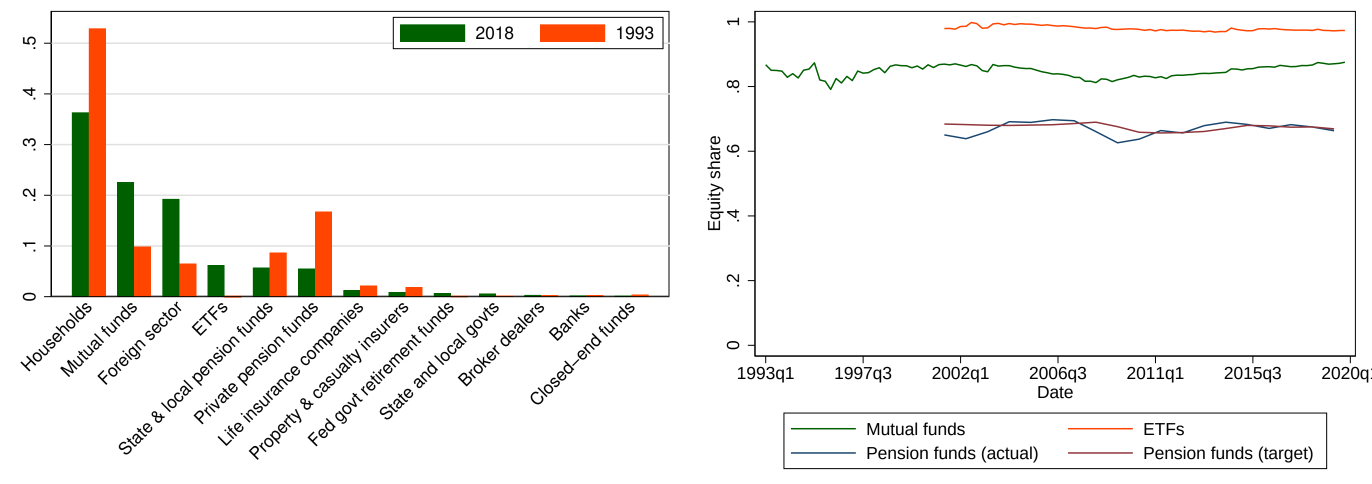
We use monthly disaggregated data on assets under management, the share invested in US equities, and flows from Morningstar for mutual funds and ETFs that are domiciled in the US and that have the US dollar as the base currency. We select the funds in Morningstar's US category groups "US Equity," "Sector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

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

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

[6] Data of different vintages can be downloaded from this website.
[7] We remove fund of funds in our analysis to avoid double counting.
[8] We omit a small number of fund-quarters in which the US equity share exceeds 300% or is lower than -300%, as these may be data errors.

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



is from 1999 to 2019.

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

Lastly, we use survey expectations of returns from Gallup, as also used by Greenwood and Shleifer (2014), who use the fraction of investors who are bullish (optimistic or very optimistic) minus the fraction of investors who are bearish. We update their data, which starts in 1996.Q4, to 2018.Q4, and the resulting series has some gaps.

2.2 Institutions often have a quite stable equity share

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

For some of these sectors, such as mutual funds, exchange-traded funds, and pension funds, we have investor-level data on equities and fixed income holdings. In the right panel of Figure 1, we plot the equity share. We aggregate different investors in a given sector using the relative sizes of their equity portfolios as opposed to assets under management, consistently with our theory (see the discussion around 15). To appreciate the importance of this difference, consider an economy with only pure equity and pure bond funds that have the same amount of assets under management. The equity-weighted equity share equals 100% while the asset-weighted equity share equals only

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

The plot shows that equity shares are quite stable over time for broad classes of investors. This is consistent with many institutions having a rather rigid mandate to maintain a stable equity share. In the model that we introduce in Section 3, this mandate rigidity will be captured by a low elasticity (κ) of funds' asset location to the expected return on equities. In recent work, Cole et al. (2021) show that a large fraction of households[9] also have a high average equity share at 79.2% with little variation over time (the equity-weighted equity share only drops to 76.4% at the end of 2008). This stability is in part explained by the introduction of target date funds.

2.3 In search of macro arbitrageurs

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

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

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

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

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

[10] 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.

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

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

This paper is about the macro-elasticity of the market (that is, how the aggregate stock market's valuation increases if one buys \$1 worth of stock by selling \$1 worth of bonds). This is in contrast with the very large literature that studies the micro-elasticity of the market (which describes how much the relative price of two stocks changes if one buys \$1 of one, and sells \$1 of the other).

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

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

Taken together, the existing evidence in the literature suggests a micro multiplier around 1 (so, a micro elasticity around 1), and a factor or macro multiplier that is well above 1 (so, a macro elasticity below 1).

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

[12] Also, the empirical market microstructure estimates of price impact are larger than what we find: the price impact that the microstructure literature finds is a factor of about 15 (Bouchaud et al. (2018); Frazzini et al. (2018)), which may make our estimate of 5 seem moderate. Microstructure results are typically couched in a form such as "buying 2.5% of the daily volume of a stock creates a permanent price increase of 0.15%". At first glance, values in this range might appear to imply a small price impact. However, they work out to a large price impact multiplier of $M = 15$: with 250 days of trading in a year, and a 100% per year turnover, the trade in our example would represent a purchase of $\frac{2.5\%}{250} = 0.01\%$ of the market capitalization of a stock, so that the impact of 0.15% on the price results in a multiplier of 15. The interpretation of this kind of microstructure estimates requires some caution, as we discuss in Section G.5. To sum up, a microstructure estimate of 15 may have the following interpretation: in inelastic markets with a micro elasticity equal to 1, a large market-wide desired trade ("metaorder") is on average split into 15 smaller trades executed over time, by one or several institutions collectively (for example, by three funds pursuing a similar strategy, each splitting their desired position change into five smaller trades). These microstructure estimates are also themselves to be taken with caution, since identification tends to be difficult as trades are not exogenous to prices. Using high frequency data with a GIV-based identification may be a promising way to enrich identification procedures in microstructure.

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

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

How do these estimates compare to the elasticities implied by standard asset pricing models? It is well known (e.g. Petajisto (2009)) that the micro elasticity in standard models is very large, of the order of 1000 or above. This implies that the micro multiplier (the inverse of the micro elasticity) is essentially zero and "demand curves are virtually flat." Based on the estimates reported in Table 1, the models are several orders of magnitudes off in terms of the micro elasticity.

Our focus is on the macro elasticity and we compute it for various asset pricing models in Section F.4.[13] The summary is that in traditional, elastic asset pricing models the macro elasticity is around 10 to 20, leading to a multiplier around 0.1 to 0.05. As any two stocks are closer substitutes than stocks and bonds, the micro multiplier is much lower than the macro multiplier in standard asset pricing models. However, the micro multiplier as estimated in the literature (see Panel A) is already an order of magnitude larger than the macro multiplier implied by standard asset pricing models. The macro multiplier estimates are even larger, which deepens the disconnect between existing estimates and asset pricing models. A multiplier of 0.05 implies that if a sovereign wealth fund, for instance, were to buy 10% of the US aggregate stock market, prices would rise by just 50bp.

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

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

We quantify this via two surveys. We provide a detailed discussion in Section E and summarize the main insights here. We conducted a first survey by putting out a request via Twitter (using the #econtwitter tag) to complete an online survey. In addition, we asked participants of an online seminar at VirtualFinance.org to complete the same survey – this latter audience being naturally more representative of the population of academic researchers in finance. Both surveys were conducted before the paper was available online and before the seminar was conducted. We received 192 responses for the Twitter survey and 102 responses for the survey connected to the finance seminar.

The survey question was the following: "If a fund buys \$1 billion worth of US equities (permanently; it sells bonds to finance that position), slowly over a quarter, how much does the aggregate market value of equities change?" The answer given in this paper is M times a billion, where M is the macro multiplier, which we estimate to be around $M = 5$. In both surveys, the median answer was $M = 0$: surveyed economists, logically enough, rely on the traditional asset pricing model in which prices are unperturbed by flows. The median positive answer was $M = 0.01$.[14] Hence, surveyed economists' views are in line with the traditional model, but far from the estimates reported in the empirical literature, and the new estimates we provide.

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

3 The Inelastic Markets Hypothesis: Theory

We now provide a model that we think is more realistic to think concretely about the determinants of stock demand, and about how flows impact prices. It is highly stylized, but will be useful to think about the determinants of elasticity (both conceptually and in terms of calibration) and to guide empirical work. We start with a two-period version, and then proceed to an infinite-horizon variant.

3.1 Two-period model

There is a representative stock in fixed supply of Q shares, with an endogenous price P. The economy lasts for two periods $t = 0, 1$. The dividend D is paid at time 1. We call $\pi = \frac{D^e}{P} - 1 - r_f$ the equity premium (with $D^e := \mathbb{E}[D]$ the expected dividend at time 0 and r_f the risk-free rate), $\bar{\pi}$ the average equity premium, and $\hat{\pi} := \pi - \bar{\pi}$ the deviation of the equity premium from its average. There is also a riskless bond with time-0 price equal to 1 (we endogenize the risk-free rate in Section 5).[15]

A representative consumer invests into stocks and bonds via I institutions or funds.[16] We call W_i fund i's wealth (or equivalently assets under management) and Q_i the number of stock market

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

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

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

shares it holds. Therefore the fraction of fund i's wealth invested in equities is $\frac{PQ_i}{W_i}$. We assume that fund i's demand for stocks is given by a mandate, saying that it should have a fraction invested in equities equal to:[17]

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

while the rest is in the riskless bond. In the simplest case, $\kappa_i = 0$, fund i has a fixed mandate to invest a fraction $\theta_i \geq 0$ of its wealth in equities. When $\kappa_i > 0$ the fund allocates more in equities when they have higher expected excess returns (hence, κ_i indexes how contrarian or forward-looking the fund is). This demand function appears sensible, and could be micro-founded along many lines – but to go straight to the effects we are interested in, we take it as an exogenous mandate.[18,19,20] We use the index $i = 0$ for a special fund, a "pure bond fund" that only holds bonds (so, its θ_i and κ_i are 0).

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

The elasticity of demand for stocks of a fund We use bars to denote values at time $t = 0^-$, before any shocks. At that time 0^-, fund i has wealth $\bar{W}_i$, and holds $\bar{Q}_i$ shares. We assume that before the shocks, equities have an equity premium $\bar{\pi}$, so that the dividend-price ratio is at its corresponding value, $\delta = \frac{\bar{D}^e}{\bar{P}}$, where $\bar{P}$, $\bar{D}^e$ are the baseline values for the stock's price and the expected dividend.

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

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

The next proposition gives the change in demand by fund i. Its proof is in Appendix A. We perform the analysis for small disturbances f_i, d, and hence small p, q_i, here and throughout the paper.[21]

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

[18] This fund's mandate can be viewed as a stand-in for other frictions such as inertia or a rule of thumb that a behavioral household might follow for its stock allocation. As a result, the institutionalization of the market does not necessarily result in more inelasticity as it depends on how households manage their own portfolios. Parker et al. (2020) argue that the growth of target date funds made the market more elastic. In the our notation, target date funds have $\kappa_i = 0$.

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

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

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

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

$$q_i = -\zeta_i p + \kappa_i \delta d + f_i, \tag{3}$$

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

$$\zeta_i = 1 - \theta_i + \kappa_i \delta. \tag{4}$$

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

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

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

$$x_S := \sum_i S_i x_i. \tag{6}$$

So, the aggregate demand change is:

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

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

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

Proposition 2 sums this up.

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

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

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

$$\theta = \theta_S, \qquad \kappa = \kappa_S, \qquad \zeta = \zeta_S, \qquad f = f_S. \tag{8}$$

In particular, ζ is the macro elasticity of demand:

$$\zeta = 1 - \theta + \kappa \delta. \tag{9}$$

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

The "aggregate flow into equities" is non-zero even though "for every buyer there is a seller" The equity-share weighted flow $f_S = \sum_i S_i f_i$ in (8) can also be expressed as[22]

$$f_S = \frac{\sum_i \theta_i \Delta F_i}{\bar{W}^{\mathcal{E}}}, \tag{10}$$

i.e. as the sum of the dollar inflows ΔF_i into each fund i, times the marginal propensity of fund i to invest in equities, θ_i, as a fraction of the the baseline value of aggregate equities $W^{\mathcal{E}} = Q\bar{P}$.[23] At the same time the net total flow is 0, $\sum_i \Delta F_i = 0$, as one bond removed from one fund goes to another fund, and the net amount of equities purchased is 0, $\sum_i \Delta Q_i = 0$, as the net amount of shares is constant:[24]

$$\sum_i \Delta F_i = 0, \qquad \sum_i \Delta Q_i = 0. \tag{11}$$

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

The impact of flows on the aggregate price We now analyze what happens after the aggregate inflow f_S in equities. We assume from now on that $\zeta > 0$. As the supply of shares does not change, we must have $q = 0$ in the equilibrium after the flow shock. Given (7), we have $0 = q = -\zeta p + f$, and the price change must be $p = \frac{f}{\zeta}$. Proposition 3 summarizes this.[25]

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

$$p = \frac{f}{\zeta}, \tag{12}$$

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

This illustrates that flows can have large price impacts if the price elasticity of demand ζ is sufficiently low, and shows the key role of this price elasticity, which is the center of this paper.[26]

An undergraduate example To think through the economics of Proposition 3, we found the following simple, undergraduate-level example useful. Suppose that there are just two funds: the pure bond fund and the representative mixed fund, which always holds 80% in equities (the magnitude suggested by Figure 1). Then, $\theta = 0.8$, $\kappa = 0$, so that $\zeta = 1 - \theta = 0.2$ and $\frac{1}{\zeta} = 5$. Then an extra 1% inflow into the stock market increases the total market valuation by 5%.

It is instructive to think through the logic of this example. Suppose that the representative mixed fund starts with \$80 in stocks (of which there are 80 shares, worth \$1 each) and \$20 in

[22]Indeed, as $\theta_i = \frac{\bar{W}_i^{\mathcal{E}}}{\bar{W}_i}$,we have $f_S = \sum_i S_i f_i = \sum_i \frac{\bar{W}_i^{\mathcal{E}}}{\bar{W}^{\mathcal{E}}} \frac{\Delta F_i}{\bar{W}_i} = \frac{1}{\bar{W}^{\mathcal{E}}} \sum_i \theta_i \Delta F_i^{\mathcal{E}}$.

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

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

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

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

bonds. There are also $B worth of bonds outstanding. Suppose now that an outside investor sells $1 of bonds from the pure bond fund (he had $B − $20 in the pure bond fund, and now he has $B − $21), and invests this $1 into the mixed fund. In terms of "direct impact", there is a $0.8 extra demand for the stock (equal to 1% of the stock market valuation), and $0.2 for the bonds. But that is before market equilibrium forces kick in.

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

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

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

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

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

$$f := f_S + f_C, \tag{14}$$

which is equal to the size-weighted total flow in the funds, f_S, plus share repurchases (as a fraction of the market value of equities). In short, on top of the traditional flows of investors into equities, we want to add share repurchases by corporations. In addition, if firms have a supply elasticity ζ_C, then the basic equilibrium is: $f_S - \zeta p = -f_C + \zeta_C p$. That is, a change in demand $f_S - \zeta p$ equals a change in supply $-f_C + \zeta_C p$. Therefore $p = \frac{f_S + f_C}{\zeta + \zeta_C}$, so that the effective market elasticity is $\zeta + \zeta_C$. In much of the paper, we assume that the supply of shares is inelastic, $\zeta_C = 0$, which will prove to be a good approximation.

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

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

which can also be expressed as $\theta_W = \frac{\sum_i W_i \theta_i}{\sum_i W_i}$. The other one is the equity-holdings weighted equity share defined earlier, $\theta_S = \frac{\sum_i W_i^{\mathcal{E}} \theta_i}{\sum_i W_i^{\mathcal{E}}}$, where $W_i^{\mathcal{E}}$ was the equity holding of fund i. The former share (θ_W) is directly available in aggregated data, while the latter (θ_S) is what matters for the macro elasticity. They are different, and indeed $\theta_S > \theta_W$.[27] This makes the disaggregation issues potentially non-trivial, and will require some care in the empirical part.

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

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

3.2 Infinite horizon model

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

We consider the case where there is a pure bond fund and "representative mixed fund" trading stocks and bonds. This allows us to zoom in on the core economics: an economy with several funds can be represented via a single mixed fund to the leading order, as in Proposition 2.[28] The representative mixed fund has a mandate: the fraction invested in equities, $\frac{P_t Q_t}{W_t}$, should be

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

where as before $\hat{\pi}_t := \pi_t - \bar{\pi}$ is the deviation of the equity premium from its average, and we allow for additional demand shocks, ν_t. These can be thought of as shocks to tastes or perceptions of risk. We assume that dividends and interest rates on bonds are passed to consumers: hence, reinvesting dividends counts as an inflow.

To analyze this economy, it is useful to linearize it. This needs to be done around a simpler, "baseline" economy, which is on a balanced growth path with a constant equity premium $\bar{\pi}$. We call $\bar{P}_t$, $\bar{D}_t$, $\bar{W}_t$, and $\bar{Q}$ the baseline price, dividend, wealth, and quantity of shares held by the mixed fund. We assume that $\left(\bar{P}_t, \bar{D}_t, \bar{W}_t\right) = \left(\bar{P}_0, \bar{D}_0, \bar{W}_0\right) \mathcal{G}_t$: they grow with a common cumulative growth factor $\mathcal{G}_t$, such that $\frac{\mathcal{G}_{t+1}}{\mathcal{G}_t}$ follows an i.i.d. growth process with mean g. As the equity premium is always $\bar{\pi}$ in the baseline economy, $r_f + \bar{\pi} - g = (1+g)\,\delta$, with $\frac{\bar{P}_t \bar{Q}}{\bar{W}_t} = \theta$ and $\frac{\bar{D}_t}{\bar{P}_t} = \delta$.[29] At the same time, the bond holdings of the mixed fund are $\bar{B}_0 + \bar{F}_t$, where $\bar{F}_t$ is the cumulative dollar inflow since time 0 (so $\bar{F}_0 = 0$): the only "new" bonds that the representative mixed fund has must come from inflows, like in the undergraduate model above. They should also represent a fraction $1-\theta$ of the wealth of the fund, so that we have: $\bar{B}_0 + \bar{F}_t = \frac{1-\theta}{\theta} \bar{P}_t Q$. This means that $\bar{F}_t = \frac{1-\theta}{\theta}\left(\bar{P}_t - \bar{P}_0\right)\bar{Q}$. This is the flow consistent with a balanced growth path in the rational economy.

We call p_t, w_t, d_t, q_t the deviations from the baseline, so that $d_t = \frac{D_t}{\bar{D}_t} - 1$, $p_t = \frac{P_t}{\bar{P}_t} - 1$, $w_t = \frac{W_t}{\bar{W}_t} - 1$, and $q_t = \frac{Q_t}{\bar{Q}} - 1$. We define the flow f_t as the scaled cumulative inflow in excess of the baseline:[30]

$$f_t = \frac{F_t - \bar{F}_t}{\bar{W}_t}. \tag{17}$$

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

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

[28] This is detailed in Appendix G.7.

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

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

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

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

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

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

As the total number of shares is constant, the equilibrium condition is given by $q_t = 0$. This yields the stock price as follows (the proof is in Appendix A).

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

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

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

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

The deviation of the equity premium from its average is:

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

We next analyze the economics of Proposition 5. The classical (or undergraduate) "efficient markets" benchmark, where the risk premium is kept constant by very strong arbitrage forces, corresponds to $\kappa = \infty$, so that $\zeta = \infty$ and $\rho = \delta$.[31]

In (20), the price discounts future dividends at a rate $\rho \geq \delta$ given in (21). So, *the market is more myopic (higher* ρ*) when it is less sensitive to the equity premium (lower* κ*) and when the mixed fund has a lower equity share (lower* θ*).*[32] It makes good sense that a lower sensitivity to the equity premium makes the market less reactive to the future, hence more myopic.[33,34] In the rest of this section, we set $\nu_t = 0$; the general case simply comes from replacing f_t by $f_t + \nu_t$.

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

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

[33] The intuition for the sign of the impact of θ on ρ is as follows: the extra term $\frac{1-\theta}{\kappa}$ in $\rho = \delta + \frac{1-\theta}{\kappa}$ is the ratio of the "present looking" (myopic) demand elasticity $1 - \theta$ to the "forward looking" elasticity κ. Hence a higher θ leads to a less myopic demand. This myopia in (20) generates momentum: because the market is myopic (by (20)), dividend news are only slowly incorporated into the price.

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

A permanent inflow has a permanent effect on the price and future expected returns of equities Suppose that at time 0 there is an inflow f_0 that does not mean-revert. Then, the impact on the price at time $t \geq 0$ is (via (20), with $\mathbb{E}_0[f_\tau] = f_0$):

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

So, the "price impact" is permanent. As a result, the equity premium is permanently lower, $\mathbb{E}_0[\hat{\pi}_t] = -\delta \frac{f_0}{\zeta}$ (see (18)) This is simply because, if the equity demand has permanently increased, equity prices should be permanently higher.[35]

Quantitatively, if prices increase by 10% due to uninformed flows, the per annum expected excess return falls by a mere 0.3% (indeed, assuming a dividend yield of 3%, $\hat{\pi} = -\delta p = -3\% \times 0.1 = 0.3\%$). This is a vivid reminder that the absence of detectable market timing strategies tells us little about market efficiency (Shiller (1984)). Similarly, Black (1986) famously argued that the aggregate stock market can be mispriced by as much as a factor of two; in our model, if this is due to a permanent inflow, that would lead to a 2% change in the expected excess return,[36] which is less than a single standard error deviation of the expected excess return estimate if one were to use 30 years of data.

The impact of a mean-reverting flow Suppose now that at time 0 there is an inflow f_0 that mean-reverts at a rate $\phi_f \in [0, 1]$, so that the cumulative flow is $\mathbb{E}_0[f_\tau] = (1 - \phi_f)^\tau f_0$. Then, if there are no further disturbances, the impact on the time-t price is $p_t = \frac{f_t}{\zeta + \kappa\phi_f}$ (see (20)), implying

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

and the change in the equity premium is $\mathbb{E}_0[\hat{\pi}_t] = -\frac{\delta + \phi_f}{\zeta + \kappa\phi_f}(1 - \phi_f)^t f_0$ (see (22)). Hence, an inflow that has faster mean reversion leads to a smaller change in the price of equities (compared to a permanent inflow), but a *larger* change in their equity premium on impact (indeed, $\frac{\delta + \phi_f}{\zeta + \kappa\phi_f}$ is increasing in ϕ_f). Those effects dissipate as the inflow mean-reverts, at a rate ϕ_f.

Predictable future inflows or changes in fundamentals create predictable price drifts Suppose that it is announced at time 0 that a permanent inflow f_T will happen at time $T > 0$. The price impact for $t \in [0, T]$ is $p_t = \frac{1}{(1+\rho)^{T-t}} \frac{f_T}{\zeta}$ (see (20), using $f_\tau = 1_{\tau \geq T} f_T$), so that after the initial jump, the price gradually drifts upward (assuming for concreteness that the inflow is positive). Hence, the risk premium is elevated by $\hat{\pi}_t = \frac{1-\theta}{\kappa} p_t$ (for $t \in [0, T)$, see (22)), and more elevated as one nears the inflow. After the inflow, though, we are back to the case of a permanently elevated price and permanently lower equity premium ($p_t = \frac{f_T}{\zeta}$ and $\hat{\pi}_t = -\delta \frac{f_T}{\zeta}$ for $t \geq T$). The same price drift before the shock happens for a predictable increase in future fundamentals such as dividends.

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

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

[35] In a Kyle (1985) model, flows change prices, like in our model; but they do not on change the equity premium (on average), which is a crucial difference with our model. Section G.5 details the link with the Kyle model.

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

with $\mathbb{E}_{t-1}\left[\varepsilon_t^f\right] = 0$. Then, a high inflow increases equity prices and hence lowers the equity premium, in the following precise manner:[37]

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

Calibration We want to understand how a macro price impact of $M \simeq 5$ might arise, and for this we calibrate the model. When flows are mean-reverting with speed ϕ_f, the price impact is $M = \frac{1}{\zeta^M}$, with $\zeta^M = \zeta + \kappa\phi_f = 1 - \theta + \kappa\left(\delta + \phi_f\right)$ (see (9), (24), and (26)). Some parameters are easy to estimate. We take a dividend-price ratio $\delta = \frac{D}{P} = 3.7\%$/year (we use annualized units throughout).[38] We calibrate $\phi_f = 4\%$/year to match the speed of mean-reversion of the dividend-price ratio.[39] Given the results in Figure 1, we take an equity share $\theta = 87.5\%$ (equity-holdings weighted as in θ_S).

Calibrating κ is most challenging. We perform a few thought experiments to see what we might expect κ to be. The simplest rational model of portfolio choice where $\theta_{it} = \frac{\pi_t}{\gamma_i \sigma^2}$ gives $\kappa = \frac{d \ln \theta_{it}}{d\pi_t} = \frac{1}{\bar{\pi}} = 22$, using an annual equity premium of 4.4%.[40] But, we rarely observe such large swings in investors' portfolios: the frictionless rational model predicts agents that are much too reactive, like in much of this paper, and in much of economics (Gabaix (2019)). To get a further feel for κ, suppose the equity premium increases from $\pi_t = 5\%$ to $\pi_t = 10\%$, which is a shift equal to about one to two standard deviations of its unconditional time-series variation (Cochrane (2011); Martin (2017)). A very flexible fund with an average equity share of 50% might change its equity allocation from 50% to 75%. This flexible fund would have $\kappa_i = \frac{d \ln \theta_i}{d\pi} = \frac{\ln 0.75 - \ln .5}{0.05} \simeq 8$. However, these are large swings in a fund's strategic asset allocation that are not typically observed empirically, so that they are at most valid only for very flexible investors. As many balanced funds have a fixed-share mandate and $\kappa = 0$, we hypothesize a κ_i for a typical fund with equity share of 50% equal to about 4. Moreover, a 100% equity funds needs to have $\kappa_i = 0$; more generally, the rigidity mechanically should increase with the equity share θ_i. So, we might tentatively parametrize a typical value of κ as $\kappa_i = K\left(1 - \theta_i\right)$, with $K \simeq 8$. So, we obtain $\kappa = \kappa_S = K\left(1 - \theta_S\right) \simeq 8 \times (1 - 0.88) \simeq 1$. This gives a simple microeconomic interpretation for the value $\kappa = 1$. Together, this yields $\zeta = 0.16$, and $\zeta^M = 0.2$, so that the price impact is indeed $M = \frac{1}{\zeta^M} = 5$. If the flows are extremely persistent, the subtle difference between ζ and ζ^M vanishes ($\kappa\phi_f$,which is 0.04 in the calibration, goes to 0).

4 Estimating the Aggregate Market 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 macroeco-

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

[38]Section H.1 details how to go from continuous to discrete time.

[39]We compute the dividend yield by summing dividends during the last 12 months relative to the current level of the CRSP value-weighted return index from January 1947 to December 2018. The annual autocorrelation of the log dividend yield during this sample is equal to $\rho^{OLS} = 0.91$ with OLS standard errors equal to 0.048. We then remove the Kendall (1954) bias $\frac{1+3\rho}{T}$ over our sample of $T = 72$ years, which is around $\frac{4}{72}$. Thus we calibrate $\phi_f = 1 - \rho^{OLS} - \frac{4}{72} \simeq 4\%$.

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

nomics. In the context of asset pricing, large literatures try to estimate the coefficient of relative risk aversion, the elasticity of inter-temporal substitution, and the micro-elasticity of demand, but not the macro elasticity.[41]

The key difficulty is that prices and flows are in part driven by other variables, such as macroeconomic news, so that naively regressing prices on flows or flows on prices would not yield a consistent estimate of the elasticities. Hence, we need an instrument. In this section, we provide first estimates of the macro elasticity of the US stock market using the method called Granular Instrumental Variables (GIV), which we conceived for the present paper, and lay out in Gabaix and Koijen (2020). Given the relevance of this parameter, we believe it would be valuable for future empirical asset pricing research to explore different estimation and identification strategies in estimating its value.

In Section 4.1, we provide the basic intuition behind the GIV estimator. We report the estimates in Section 4.2 using sector-level data from the Flow of Funds and in Section 4.3 using investor-level data by combining 13F filings and mutual fund flows. We also connect capital flows to macroeconomic variables and measures of beliefs to provide an initial analysis of the potential determinants of flows into the equity market.

4.1 Intuition behind the GIV estimator

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

$$X_E := \frac{1}{N}\sum_{i=1}^{N} X_i, \qquad X_S := \sum_{i=1}^{N} S_i X_i, \qquad X_\Gamma := X_S - X_E. \tag{27}$$

Suppose that we have a time series of changes in investors' equity holdings, $\Delta q_{it} = \frac{Q_{it}-Q_{i,t-1}}{Q_{i,t-1}}$, where i indexes investors as before. The estimation procedure does not require data on flows across asset classes: equity holdings suffice, which is empirically relevant as investor-level data on flows are available only for a subset of investors.[42] To fix ideas, we model Δq_{it} as (omitting constants):[43]

$$\Delta q_{it} = -\zeta \Delta p_t + f_{it}^{\nu}, \tag{28}$$

where Δp_t is the aggregate stock return, and ζ is the demand elasticity of interest — we take it as constant across sectors in this introduction, but will relax this in Section 4.3. We consider the following model for f_{it}^{ν}:

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

[41] See Table 1 for a summary of estimates in the literature.

[42] If we were to have high-quality data on capital flows, f_{it}, then we could construct another granular instrumental variable using capital flows by extracting idiosyncratic shocks to f_{it}. However, our theory implies that we need accurate data on equity and bond holdings to measure capital flows correctly, which are unavailable in the US. Fortunately, however, we can implement the GIV procedure using Δq_{it}, which does not require knowledge of holdings in other assets than equities. Also, in Section 4.3 we show how to use data on a subset of investors, in our case mutual funds, to obtain another estimate.

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

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

$$\mathbb{E}\left[u_{it}\eta_t\right] = 0. \tag{30}$$

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.

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

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

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

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

$$Z_t = u_{St} - u_{Et} = u_{\Gamma t}. \tag{31}$$

As $u_{\Gamma t}$ is a combination of idiosyncratic shocks only, it is uncorrelated with η_t, see (30). This orthogonality condition makes $Z_t = u_{\Gamma t}$ a valid instrument: it is our GIV. Furthermore, if u_{it} is homoskedastic, then $u_{\Gamma t}$ is uncorrelated with u_{Et}.[44]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

$$\Delta p_t = M Z_t + e_t, \tag{32}$$

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

$$\Delta q_{Et} = -\zeta \Delta p_t + \epsilon_t, \tag{33}$$

with Δp_t instrumented by Z_t.

Intuitively, we use the sector-specific, or idiosyncratic, demand shocks of one sector as a source of exogenous price variation to estimate the demand elasticity of another sector. Viewed this way, the GIV estimator generalizes the idea behind the index inclusion literature to estimate the

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

micro elasticity. In the index inclusion literature, a demand shock to the group of index investors (assuming the inclusion of a stock into the index is random) can be used to estimate the slope of the demand curve of the non-index investors.

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

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

$$\Delta \check{q}_{it} = \check{\lambda}_i' \eta_t + \check{u}_{it}. \tag{34}$$

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

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

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

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

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

GIV: Requirements and threats to identification For the GIV to be consistent, we need (30) to hold: the idea is that there are random "bets" or "shocks" to various fund managers, institutions and sectors, that are orthogonal to all reasonable common macro factors such as GDP, TFP, and so forth. For the GIV to be a powerful instrument, we need large idiosyncratic shocks, and a few large institutions, so that the market is "granular" in the sense that the idiosyncratic trading shocks of a few large players meaningfully affect the aggregate.[45] Fortunately, this is verified in our setting, as it is in related settings in macro (Gabaix (2011), Carvalho and Grassi (2019)), trade (Di Giovanni and Levchenko (2012)) or finance (Amiti and Weinstein (2018), Herskovic et al. (forthcoming), Galaasen et al. (2020)). Ben-David et al. (forthcoming) and Ghysels et al. (2021) study the impact of investor granularity on the cross-section of US stock returns.

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

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

When firms are elastic and flows mean-revert When firms have a supply elasticity ζ_C, the total elasticity is $\zeta + \zeta_C$, as we saw in Section 3.1. When flows mean-revert with speed ϕ_f, the measured elasticity is $\zeta + \kappa\phi_f$, as we saw in (24) and (26). Combining those two extensions, the measured price impact is

$$M = \frac{1}{\zeta^M}, \qquad \zeta^M := \zeta + \kappa\phi_f + \zeta_C. \tag{36}$$

As $\kappa\phi_f$ and ζ_C appear to be small, the difference between ζ, $\zeta + \kappa\phi_f$ and ζ^M is rather minor, and is best ignored in the first pass. Still, to be completely explicit, when we empirically measure "ζ", we actually measure a quantity that is $\zeta + \kappa\phi_f + \zeta_C$ if flows mean-revert at speed ϕ_f and firms have a supply elasticity ζ_C, and is strictly speaking ζ only when flows do not mean-revert and $\zeta_C = 0$.[46]

4.2 Elasticity estimates using sector-level data

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

$$\Delta q_{it} = \alpha_i - \zeta\Delta p_t + \lambda_i'\eta_t + u_{it}, \tag{37}$$

where we assume that the demand elasticity is the same across sectors. We relax this assumption below using 13F data. We consider the same model for the corporate sector, but allow for a different demand elasticity, ζ_C. The vector η_t includes GDP growth, a time trend,[47] and one or more latent factors, η_t^{PC}.

The results are presented in Table 2. The first column reports the estimates where we use a single principal component to isolate the idiosyncratic shocks to various sectors, in addition to a common factor on which all sectors load equally.

We estimate a multiplier of $M = 7.1$, implying that purchasing 1% of the market results in a 7.1% change in prices. The corresponding standard error is 1.9.[48] 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 (33). With one principal component, we estimate an elasticity of $\zeta = 0.13$ and with two principal components, we estimate $\zeta = 0.17$. In the next two columns, we estimate the supply elasticity ζ_C of the corporate sector. The short-run elasticity is low at $\zeta_C = 0.01$ for both one and two principal components.[49] This implies that the combined elasticity is 0.14 (with

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

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

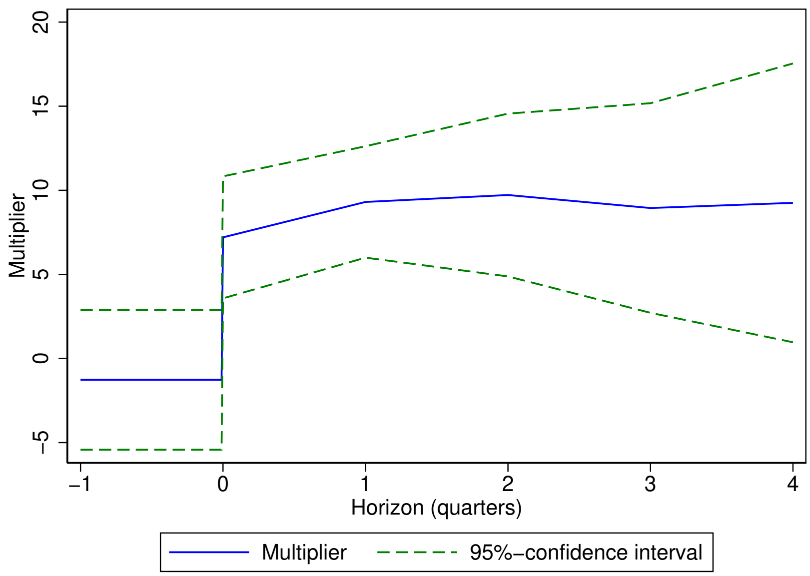
[48] We report Newey-West standard errors using the bandwidth selection as in Newey and West (1994).

[49] This small contemporaneous elasticity of the supply of shares by the corporate sector, estimated here causally by IV, is consistent with the OLS findings of Ma (2019). She finds (Table VII) that $\frac{\text{Gross equity issuance}}{\text{Assets}} = 0.01\hat{\pi}$ (plus other terms) at the quarterly frequency. Using that equity is about two thirds of assets, this leads, at the annual frequency, to $\Delta q_C = \frac{3}{2} \cdot 4 \cdot 0.01\hat{\pi} = 0.06\hat{\pi}$, so that (by (18)) $\zeta_C = \delta \times 0.06 = 0.0024$. However, these estimates do not rule out the possibility that the medium- or long-run elasticities are higher and that firms play an important role in stabilizing asset prices.

Table 2: Estimates of the macro elasticity. The first two columns report estimates of M with one and two principal components, η_t, respectively. The next two columns report the elasticity estimates, ζ, regressing the equal-weighted change in equity holdings Δq_E on the price change Δp instrumented by the GIV Z. The next two columns report the elasticity of the corporate sector, ζ_C. The final column reports the estimates of the same specification as the first column, but we omit Z_t, where $Z_t = \sum_i S_{i,t-1}\Delta\check{q}_{it}$ and $\Delta\check{q}_{it}$ defined in (63), to estimate the impact of sector-specific shocks on prices. In constructing $\Delta\check{q}_{it}$, all estimates control for quarterly GDP growth. We report the standard errors, which are corrected for autocorrelation, in parentheses. The sample is from 1993.Q1 to 2018.Q4.

	Δp	Δp	Δq_E	Δq_E	Δq_C	Δq_C	Δp
Z	7.08	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 2: Estimates of the aggregate multiplier $M = \frac{1}{\zeta}$ by horizon. The figure plots the multi-period impact of demand shocks: a demand shock of f_t at date t increases the (log) price of equities from $t-1$ to $t+h$ by Mf_t. We use the GIV for instrumentation, see (38). 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.



one principal component) or 0.18 (with two principal components). The corresponding multipliers, $M = \frac{1}{\zeta + \zeta_C}$, are $M = 7.1$ and $M = 5.9$, respectively.

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

The impact of flows at longer horizons In Figure 2, we explore how demand and flow shocks propagate across time. To this end, we extend the earlier analysis by estimating

$$p_{t+h} - p_{t-1} = a_h + M_h Z_t + c_h \eta_t^{PC,e} + d_h \Delta y_t + \epsilon_{t+h}, \tag{38}$$

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

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

Robustness We explore the robustness of our estimates along various dimensions. In the interest of space, we report and discuss the tables in Appendix D.4. In Tables D.8–D.10, we consider a variety of robustness checks related to the sample period, data construction, and implementation choices of the GIV estimator. We conclude that our results are robust to these changes in the empirical strategy with multiplier estimates ranging from 3.5 to 8.

4.3 Elasticity estimates using investor-level data

We provide an alternative estimate of the same elasticity, but now using more disaggregated, investor-level 13F and mutual fund data.[50] We use 13F data from FactSet that cover the period from 2000.Q1 to 2019.Q4 and we follow the data construction as in Koijen et al. (2019). Monthly mutual fund flows come from Morningstar from January 1993 to December 2019. We provide details in terms of the data construction in Appendix C.2. An advantage of these disaggregated data is that we can allow for heterogeneous demand elasticities across investors. To provide another estimate of the multiplier, we proceed in three steps.

First, we estimate innovations in fund flows for mutual funds. Let Δf_t be the fractional inflow into equity markets from mutual funds.[51] We estimate

$$\Delta f_t = a_0 + \sum_{l=1}^{k} a_l \Delta f_{t-l} + ct + \epsilon_{mt}^f, \tag{39}$$

at a monthly frequency (see Table D.12 in Online Appendix D.6). We also define $K = \frac{1}{1-\sum_{l=1}^{k} a_l}$, which is the cumulative flow due to shocks ϵ_{mt}^f: as per Proposition 5, what matters is the cumulative future inflow, which is $K\epsilon_{mt}^f$.[52]

It is well known that the innovations, ϵ_{mt}^f, are correlated with contemporaneous realized returns (Warther (1995); Edelen and Warner (2001); Goetzmann and Massa (2003)). We extend this literature by removing aggregate demand factors and isolating the idiosyncratic demand shocks of mutual fund investors. In addition, we show how to translate the persistence in flows to a theory-based estimate of the multiplier via K. We aggregate the monthly innovations, ϵ_{mt}^f, for $k = 3$ in each quarter and refer to these innovations as ϵ_t^f. We model

$$\epsilon_t^f = \beta_0' \eta_t + \beta_1' C_t + u_t^f,$$

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

[51] To compute the relevant measure of flows, we start from the share invested in US equities by fund i, θ_{it}, assets under management, W_{it}, and the flow ΔF_{it} as defined by Morningstar. When equity shares are missing at a monthly frequency, we fill in the equity shares using the most recent value for a given fund. We first compute $\Delta f_{it} = \frac{\Delta F_{it}}{W_{i,t-1}}$ and winsorize it at 1% and 99% in each period. We then compute $\Delta f_t = \frac{\sum_i \theta_{i,t-1} W_{i,t-1} \Delta f_{it}}{\sum_i \theta_{i,t-1} W_{i,t-1}}$, which uses equity-weighting, as warranted by the theory, see (10). We include "US equity," "sector equity," "international equity," and "allocation" funds in the analysis. Appendix C.2 provides additional details.

[52] In principle, it should be discounted at a rate ρ, but this is immaterial at the horizon of a few months that we use. See Section G.5 for details.

where η_t are common unobserved factors, C_t are common observed factors, and u_t^f are the unique shocks to fund flows.[53]

Second, we wish to estimate those common factors, η_t, to isolate the shocks that are unique to mutual fund investors. To do so, we use the 13F filings of investors outside of the mutual fund industry (e.g., pension funds, insurance companies, and so forth).[54] We consider an extension of the model in (37), where we allow for heterogeneity in demand elasticities, $\zeta_{i,t-1}$:

$$\Delta q_{it} = \alpha_i - \zeta_{i,t-1}\Delta p_t + \lambda_{i,t-1}'\eta_t + \beta_i' C_t + u_{it}.$$

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

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

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

Third, we regress returns on fund flow innovations, while controlling for common factors,

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

where $Z_t = K S_{t-1}^{MF} \epsilon_t^f$, with S_{t-1}^{MF} the share of aggregate equities held by the mutual fund sector: after controls, this is the surprise inflow unique to mutual funds. As a common observed factor, C_t, we use GDP growth. We also explore robustness to controlling for changes in volatility in C_t.

The results are summarized in Table 3. The first column presents the results with only Z_t and GDP growth, something we show for illustration but do not recommend using. The next four columns add the factors extracted from the 13F data, η_t^e, as recommended by the GIV. In the final column, we also control for the quarterly (percentage) change in volatility. Without controls other than GDP in Column 1, the multiplier estimate equals $M = 10.9$. By adding additional controls, the R-squared value increases significantly and the multiplier estimate lowers, as we would expect since demand shocks and prices are positively correlated. With four additional factors, the R-squared value equals approximately 60% and the multiplier drops to $M = 7.7$ with a standard error of 2.3. In the final column, we add changes in volatility. While these do not correlate strongly with fund flow innovations, they do correlate with returns. This suggests that other investors are negatively sensitive to volatility and this also captures a source of demand shocks. The multiplier lowers further to 7.6 and the R-squared is now 70%.[55] In Figure D.8, we also repeat the long-horizon analysis as in Figure 2. As before the impact of flow shocks on prices is persistent although the confidence interval is wide at longer horizons of one year.

In summary, we find that the multiplier estimates are quite consistent with the estimates we found using the FoF data. These estimates well above 1 are consistent with the estimates for other

[53] Flows themselves may be sensitive to prices, so that $\epsilon_t^f = -\zeta^f \Delta p_t + \beta_0' \eta_t + \beta_1' C_t + u_t^f$. In this case, if ζ^f is negative, as is the case when mutual fund investors engage in positive feedback trading, then our estimates provide a lower bound.

[54] Specifically, we omit investors outside of the mutual fund industry using the same assignment of investor types as in Koijen et al. (2019). This removes, for instance, investment advisors and mutual funds as classified by FactSet.

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

Table 3: Estimates of the macro elasticity using mutual fund and 13F data. The first five columns provide estimates of the multiplier M, which is the coefficient on $Z_t = K S_{t-1}^{MF} \epsilon_t^f$, the innovation in the cumulated inflow into mutual funds after controls. We regress returns on unexpected flows, ϵ_t^f, times the share of aggregate equities held by the mutual fund sector, S_{t-1}^{MF}, and adjusting for the fact that inflows are autocorrelated (see (39) and the surrounding definition of K). In the first column we only control for GDP growth and in the next four columns we add one to four common factors to isolate the idiosyncratic component in mutual fund flows. The common factors are extracted from 13F filings of institutions outside of the mutual fund industry. In the final column, we add the change in quarterly volatility as an additional control. We report the standard errors, which are corrected for autocorrelation, in parentheses. The sample is from 2000.Q1 to 2019.Q4.

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

countries and for style factors, see Table 1. Future research can explore other strategies to control for common demand factors to sharpen the identification.

4.4 A new measure of capital flows into the stock market

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

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

Measuring flows into the stock market We rely on the FoF data for these calculations and we refer to Appendix C for details on the data construction of the fixed income positions and flows. As (10) shows, the flow into the aggregate stock market can be measured by first computing the flow for each investor, $\Delta f_{it} = \frac{\Delta F_{it}}{W_{i,t-1}}$, and then computing the equity weighted average, $\Delta f_{St} = \sum_i \frac{\theta_{i,t-1} W_{i,t-1}}{\sum_i \theta_{i,t-1} W_{i,t-1}} \Delta f_{it}$. Unfortunately, the FoF aggregates data across many institutions and the reported flows can be mismeasured by this definition. To see this, consider the case in which some households only invest in bonds and other households only invest in equities. If we view this as a combined household, a 1% combined inflow into financial markets does not necessarily lead to a 1% increase in equity holdings as the flow may be a flow to bond funds only. With disaggregated data, such problems can be solved, but such data are unfortunately unavailable.

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

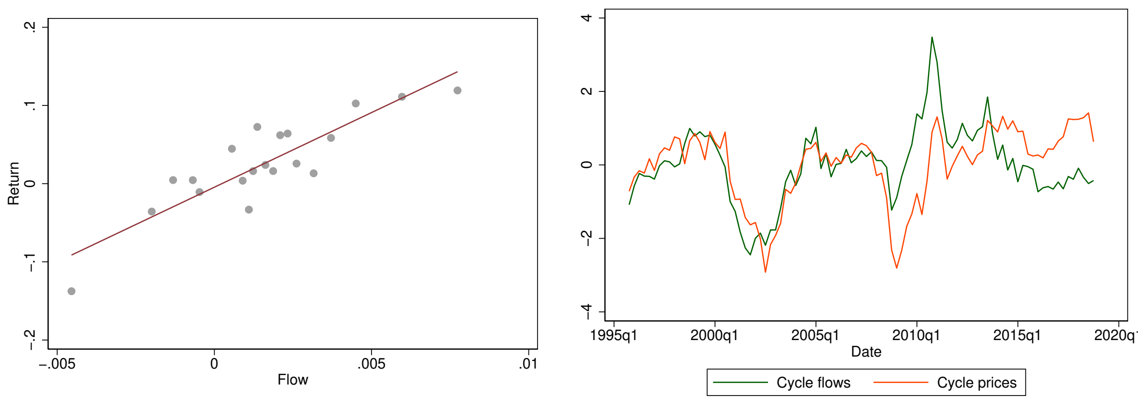
$$\Delta q_{it} = \alpha + \beta_i f_{it} + \gamma_i \Delta p_t + \delta_i \Delta y_t + \epsilon_{it}. \tag{40}$$

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

The correlation between capital flows and equity returns We relate our measure of capital flows into the stock market to returns. In the left panel of Figure 3, we show that our measure of flows is strongly correlated with returns using a binned scatter plot in the left panel. We again find that the slope is high, but we emphasize that, because of endogeneity, the slope is not a good measure of the impact of flows of the price. This is why earlier we developed an IV strategy to measure that impact.[56]

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

Figure 3: Capital flows into the stock market and price changes. We plot the aggregate flow into the stock market, using the screened flows, f^*_{jt}, $f^*_{St} = \sum_{j=0}^{N} S_j f^*_{jt}$, versus the return on the aggregate stock market in the left panel used a binned scatter plot. In the right panel, we construct a cumulative (log) return index and compute cumulative flows. We extract the cyclical component using the methodology developed in Hamilton (2018). We standardize both measures over the full sample to be able to plot them in the same figure. The sample for both figures is from 1993.Q1 to 2018.Q4.



We can also illustrate the strong co-movement between flows and prices at lower frequencies. In particular, we construct a cumulative (log) return index and compute cumulative flows. We then extract the cyclical component using the methodology developed in Hamilton (2018). We standardize both measures, by removing the time-series mean and dividing them by their standard deviations, over the full sample to be able to plot them in the same figure. These are shown in the right panel of Figure 3. Consistent with the high-frequency co-movement that we uncover in the left panel of Figure 3, we find that prices and flows co-move at a business cycle frequency. We re-emphasize once again that these are merely correlations and it may be the case, for example, that they reflect positive feedback trading by investors (Cutler et al. (1990), Shleifer (2000)).

Relating flows to shocks to GDP and to return expectations To conclude this initial exploration of capital flows into equity markets, we relate flows to shocks to economic activity and survey expectations of returns. We use GDP growth as our measure of economic activity, as before. For return expectations, we use the survey from Gallup. The data are described in more detail in Appendix C. Gallup has several missing observations and only starts in 1996.Q4. We only use data for all series when they are non-missing, which gives us 79 quarterly observations. To obtain innovations, we estimate an AR(1) model for each of the series (except returns). We standardize each of the innovation series, by removing the time-series mean and dividing them by their standard deviations, to simplify the interpretations of the regressions.

The results are presented in Table 4. In the first three columns, we relate capital flows to survey expectations and economic growth. We find that flows and survey expectations are strongly correlated, confirming Greenwood and Shleifer (2014) using a more comprehensive measure of capital flows. A one standard deviation increase in survey expectations of future returns is associated with a 0.48 standard deviation increase in capital flows.

This finding may point to a resolution of a recent challenge posed to the beliefs literature by

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

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

Giglio et al. (2021a). In particular, they find that although survey expectations of returns are volatile, the pass-through to actions (that is, portfolio rebalancing) is low. One possibility is that the strong correlation between innovations to beliefs and prices (which equals 61% in our sample) arises even though the pass-through is low, but small flows into inelastic markets lead to large price effects.

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

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

5 General Equilibrium with Inelastic Markets

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

5.1 Setup

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

The endowment Y_t follows a proportional growth process, with an i.i.d. lognormal growth rate G_t: $\frac{Y_t}{Y_{t-1}} = G_t = e^{g+\varepsilon_t^y - \frac{1}{2}\sigma_y^2}$, with $\varepsilon_{t+1}^y \sim \mathcal{N}\left(0, \sigma_y^2\right)$. Utility is $\sum_t \beta^t u\left(C_t\right)$ with $u\left(C\right) = \frac{C^{1-\gamma}}{1-\gamma}$. Because empirically dividend growth and GDP growth are not very correlated, we model that GDP Y_t is divided as $Y_t = \mathcal{D}_t + \Omega_t$ into an aggregate dividend $\mathcal{D}_t$ and a residual Ω_t, where the dividend stream has i.i.d. lognormal growth, $\frac{\mathcal{D}_t}{\mathcal{D}_{t-1}} = G_t^D = e^{g+\varepsilon_t^D - \frac{1}{2}\sigma_D^2}$, so that the balanced growth path specified in Section 3.2 has a cumulative growth factor $\mathcal{G}_t = G_t^D \dots G_1^D$. The "residual" Ω_t can be thought of as a combination of wages, entrepreneurial income, and so forth (and indeed it is the vast majority of GDP).[57] The representative firm raises capital entirely through equity, and passes the endowment stream as a per-share dividend $D_t = \frac{\mathcal{D}_t}{Q}$, where Q is the number of shares of equities supplied by the corporate sector, which is an unimportant constant in this baseline model without share buybacks and issuances. Bonds are in zero net supply.[58] We write the price of equities as $P_t = \frac{D_t}{\delta} e^{p_t}$, where δ is the average dividend-price ratio and p_t is the deviation of the price from the baseline $p_t = 0$. Those quantities are all endogenous.

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

$$\theta_t = \theta \exp\left(-\kappa^D p_t + \kappa \mathbb{E}_t\left[\Delta p_{t+1}\right]\right), \tag{41}$$

which is the same as before in (1), to the leading order (in terms of deviations from the steady state), with $\kappa^D = \kappa\delta$. The formulation here is slightly more general.[59]

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

[57] Formally, it could become negative, as in Campbell and Cochrane (1999), though this is a very low probability event in our calibration. Then, the interpretation is that of a residual liability. In addition, it would be easy to keep $\mathcal{D}_t/Y_t$ stationary, at the cost of having it as one more state variable, reverting to its mean.

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

[59] 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.

of household h entails:[60]

$$Q_t^{B,h} + D_t^h + \Omega_t^h = C_t^h + \Delta F_t^h + \frac{Q_{t+1}^{B,h}}{R_{f,t}}. \tag{42}$$

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

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

The rational consumer part of the household chooses consumption (but not equity shares) to maximize lifetime utility, subject to the dynamic budget constraint for bonds (42). She takes the actions of the investor as given.[61] As she is rational, she satisfies the Euler equation for bonds:

$$\mathbb{E}_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{ft} \right] = 1, \tag{43}$$

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

The behavioral investor part of the household is influenced by b_t, a behavioral disturbance. It is a simple stand-in for noise in institutions, beliefs, tastes, fears, and so on. We assume that the investor trades (between stocks and bonds) with a form of "narrow framing" objective function (as in Barberis et al. (2006)). He seeks to maximize $\mathbb{E}_t \left[V^p \left(W_{t+1} \right) \right]$ with $V^p(W) = \frac{W^{1-\gamma}-1}{1-\gamma}$ a proxy value function. Specifically, when $b_t = 0$, he chooses his allocation $\bar{\theta}^M$ in the mixed fund as:

$$\bar{\theta}^M = \underset{\theta^M}{\operatorname{argmax}} \, \mathbb{E} \left[V^p \left(\left(1 - \theta^M \right) R_{ft} + \theta^M R_{M,t+1} \right) | b_t = 0 \right], \tag{44}$$

where $R_{M,t+1}$ is the stochastic rate of return of the mixed fund. This choice of a "narrow framing" benchmark is opposed to the fully rational value function, which would have all the Merton-style hedging demand terms, and would lead to the consumption CAPM holding on average: in particular, the equity premium would be too small, as in the equity premium puzzle (at $\bar{\pi} = \gamma cov \left(\varepsilon_t^D, \varepsilon_t^y \right)$). Instead, the above formulation with narrow framing will lead to a high equity premium $\bar{\pi} = \gamma \sigma_r^2$, where the σ_r^2 is the volatility of the stock market, which is affected by flow shocks.[62]

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

$$\Delta F_t = \frac{1-\theta}{\theta \delta} \Delta \mathcal{D}_t + \frac{1}{\delta} \Delta \left(b_t \mathcal{D}_t \right), \tag{45}$$

---

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

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

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

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

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

$$f_t = \kappa^H \hat{\pi}_t^H, \tag{46}$$

with κ^H the sensitivity to the risk premium, and to a behavioral deviation $b_t = \frac{f_t}{\theta}$. Using the empirical findings of Giglio et al. (2021a), we estimate that $\kappa^H \simeq 2$, a value that we rationalize by calibrating it in terms of other behavioral parameters. This estimate is in contrast with a rational model, which would imply $\kappa^H = \frac{1}{\bar{\pi}} \simeq 22$, a very large pass-through from beliefs to portfolio shares. A low value of κ^H means that people have "bold forecasts" (excess variations in the perceived equity premium) but make "timid choices" (small flows), very much as in Kahneman and Lovallo (1993).[63]

This type of model can be also made to match the perspective in Bordalo et al. (2020), in which all variation in prices, flows, and the perceived risk premium $\hat{\pi}_t^H$ comes from changes in the long-term growth forecast g_t (all in deviations from a trend), in a way still governed by (46): Section G.9 provides details and a calibration. One could image a richer model for the perceived risk premium $\hat{\pi}_t^H$, e.g. with extrapolative beliefs based on realized returns or growth rates. One could then work out the implications for flows (via (46)) and prices (via Proposition (5)).

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

We finally formally define the equilibrium.

Definition 1. The state vector is $Z_t = (Y_t, \mathcal{D}_t, \mathcal{D}_{t-1}, b_t)$. An equilibrium comprises the following functions: the stock price $P(Z)$, the interest rate $R_f(Z)$, and the consumption and asset allocation $C(Z)$, $B(Z)$, such that the mixed fund's allocation $\theta(P, Z)$ follows its mandate, and: (i) the consumer follows the consumption policy $C(Z)$, which maximizes utility subject to the above constraints; (ii) the investor follows the behavioral policy (45), where the average allocation in the mixed fund is given by (44), so that it is quasi-rational with narrow framing on average, but with disturbance b_t; (iii) the mixed fund's demand for stocks $Q(Z)$ follows its mandate (41); (iv) the consumption market clears, $C(Z) = Y(Z)$; and (v) the equity market clears, $Q(Z) = Q$.

5.2 Model solution

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

$$f_t = \theta b_t. \tag{47}$$

This holds for any process b_t. Now, we specialize to the case where b_t follows an AR(1) with speed of mean-reversion ϕ_f. Then, so does f_t, so that we are in the "simple benchmark" case of (25)-(26), and

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

now with an endogenous interest rate and unconditional equity premium. This AR(1) assumption is just a placeholder for richer behavioral assumptions, for example driven by time-varying beliefs (as in Caballero and Simsek (2019), Bordalo et al. (2020)), positive or negative feedback trading rules, and so on. We defer to future research for richer, empirically-grounded models of the "behavioral deviation" b_t, and hence of the flows. The limited goal of this framework is to have a simple model of the *impact* of the flows in general equilibrium, which can be fully solved and which lends itself to a number of variants. Importantly, it relies on observable flows.

Proposition 6. *The solution of the economy obtains in closed form as follows, taking the limit of small time intervals and only the first order terms in f_t. The market elasticity ζ and the "macro market effective discount rate" ρ (see Proposition 5) are:*

$$\zeta = 1 - \theta + \kappa^D, \qquad \rho = \frac{\zeta}{\kappa}. \tag{48}$$

The price of equities is:

$$P_t = \frac{D_t}{\delta} e^{p_t}, \tag{49}$$

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

$$p_t = b_f^p f_t, \qquad b_f^p = \frac{1}{\zeta + \kappa \phi_f}. \tag{50}$$

Hence the variance of stock market returns is

$$\sigma_r^2 = var\left(\varepsilon_t^D + b_f^p \varepsilon_t^f\right), \tag{51}$$

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

$$\bar{\pi} = \gamma \sigma_r^2. \tag{52}$$

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

$$\pi_t = \bar{\pi} + b_f^\pi f_t, \qquad b_f^\pi = -\left(\delta + \phi_f\right) b_f^p. \tag{53}$$

Finally, the interest rate is constant, and given by the consumption Euler equation (43)*:*

$$r_f = -\ln \beta + \gamma g - \gamma\left(\gamma + 1\right) \frac{\sigma_y^2}{2}. \tag{54}$$

Hence, we have a fairly traditional economy, except that, crucially, prices and risk premia are now driven by flows and flow risk, in addition to fundamentals, and that markets are inelastic. Hence, the equity premium is time-varying (because of flows), and on average higher than in the consumption CAPM (because it reacts to flow risk, not just fundamental risk, and because the narrow framing makes the investor react to the variance of equity returns, rather than their covariance with consumption), as given in (52).

5.3 Pricing kernel consistent with flow-based pricing

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

$$\mathcal{M}_{t+1} = \exp(-r_f - \pi_t \frac{\varepsilon_{t+1}^D}{\sigma_D^2} + \xi_t), \qquad \pi_t = \bar{\pi} + b_f^{\pi} f_t, \tag{55}$$

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

This "flow-based" pricing kernel is an alternative to the consumption-based kernel of Lucas (1978). The core economics is in how flows affect prices, and the pricing kernel (55) just reflects that. The flow f_t modifies the price P_t according to Proposition 6 and also the pricing kernel $\mathcal{M}_{t+1}$, in such a way that $P_t = \mathbb{E}_t\left[\mathcal{M}_{t+1}\left(D_{t+1} + P_{t+1}\right)\right]$ holds. The pricing kernel is in a sense a symptom rather than a cause in that market.

To sum up, the flow-based SDF (55) reacts to flows, and prices equities and bonds:

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

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

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

5.4 Calibration of the general equilibrium model

We now calibrate the general equilibrium model. This extends the calibration of Section 3.2, which is natural as the general equilibrium model is an extension of the basic infinite horizon model. We use the parameter values given in Table 5, which are all presented in annualized terms for clarity. We provide a summary discussion of our parameter choices here, leaving some details to Section H. Risk aversion is moderate, at $\gamma = 2$. The macroeconomic parameter values are standard, except for the pure rate of time preference.[64] We set a speed of mean reversion of the behavioral disturbance of $\phi_b = 4\%$/year, which induces the same speed of mean reversion for flows f_t and for the P/D ratio. Likewise, we choose its standard deviation to generate the requisite volatility of flows. For parsimony, we assume zero correlation between flow shocks and dividend shocks.

Table 6 shows the resulting moments implied by the model. It verifies that we match all the "classic" moments, for instance the risk-free rate, the average equity premium, and the volatility of

[64] To get a small risk-free rate of 1% (and only for this reason), we need to make the agents very patient, so that $\beta > 1$. Indeed, this comes from the Ramsey equation (54), which is $r_f \simeq -\ln\beta + \gamma g$ (neglecting precautionary effects, which are very small in our calibration) with $\gamma g = 4\%$. We share this issue with the overwhelming majority of the macroeconomics literature: if we normalized the average growth rate to zero, like most of the macroeconomics literature, we would not have this difficulty. It would be easy to amend that, for example by adding a small probability of a disaster risk or by using Epstein-Zin preferences. We do not do that, because we do not wish to complicate the model.

Table 5: Parameter values used in the calibration

Variable	Value
Growth rate of endowment and dividend	$g = 2\%$
Std. dev. of endowment growth	$\sigma_y = 0.8\%$
Std. dev. of dividend growth	$\sigma_D = 5\%$
Mixed fund's equity share	$\theta = 87.5\%$
Mixed fund's sensitivity to risk premium	$\kappa = 1$
Speed of mean-reversion rate of behavioral disturbance	$\phi_b = 4\%$
Std. dev. of innovations to behavioral disturbance	$\sigma_b = 3.3\%$
Time preference	$\beta = 1.03$
Risk aversion	$\gamma = 2$

Notes. Values are annualized.

Table 6: Moments generated by the calibration

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

Notes. Values are annualized.

Table 7: Some stock market moments and predictive regressions

(a) Stock market moments

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

(b) Predictive regressions

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

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

stock returns. We see that the model features a large "excess volatility": the flow shocks (with their 2.8% annual standard deviation) account for almost 90% of the variance of stock returns. It may be surprising that we can match the equity premium without any of the "modern" asset pricing ingredients, such as a very high risk aversion or disaster risk. The reason is that the preferences of our behavioral investors feature "narrow framing", which leads to an average risk premium given by $\bar{\pi} = \gamma \sigma_r^2$.

Table 7 shows more moments specific to the stock market. We broadly match the volatility of the log P/D ratio, its speed of mean reversion, and the predictive power of forecasting regressions with that P/D ratio.

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

6 Government Policy and Corporate Finance in Inelastic Markets

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

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

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

This is what a number of central banks have done. In August 1998, the Hong Kong government, when it was under a speculative attack, bought 6% of the Hong Kong stock market: this resulted in a 24% abnormal return, which was not reversed in the following eight weeks (Bhanot and Kadapakkam (2006)). This effect is not entirely well-identified, but is consistent with a large price impact multiplier $\frac{1}{\zeta}$, around 4. Likewise, the Bank of Japan owned 5% of the Japanese stock market in March 2018 (Charoenwong et al. (2020)) and the Chinese "national team" (a government outfit) owned a similar 5% of Chinese stocks in early 2020.[67] In inelastic markets, this may have a large price impact.[68] Those government purchases of equities offer a potentially attractive government policy, as they increase market values and lower the cost of capital for firms, and relax credit constraints. So, they might increase hiring and real investments by firms, and GDP. We think this is an interesting direction for future research.[69]

6.2 Corporate finance in inelastic markets

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

In a frictionless model, this does not affect the firms' values, as per Modigliani-Miller. In an inelastic model, it should now be clear that buybacks can increase the aggregate value of equities. How much depends on the rationality of households, as we now detail. For clarity and brevity, we focus on the two-period model (the same economics holds with an infinite horizon, but the expressions are more complicated; see Section G.11). At time 0, we imagine the representative firm buys back a fraction b of the equity shares, where b is small (so that the new number of shares is $Q_0' = (1-b)\, Q_0$). The buyback is financed by a fall in the time-0 dividend, so the total dividend payout falls from $\mathcal{D}_0$ to $\mathcal{D}_0' = \mathcal{D}_0 - P_0 Q_0 b$, where P_0 is the ex-dividend price, and $P_0 Q_0 b$ is used to finance the share buyback.

---

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

[66] If the government buys it for just T periods, the impact is $p = \left(1 - \frac{1}{(1+\rho)^T}\right) \frac{f^G}{\zeta}$. Set $f_t = f^G 1_{0 \le t < T}$ in (20). With the above calibration, this can be a moderate dampening if T is large enough.

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

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

[69] Brunnermeier et al. (2020) caution about potentially adverse effect if the government's purchases might become too central.

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

Proposition 7. (Impact of share buybacks in a two-period model) *Suppose that, at time 0, corporations buy back a fraction b of shares, lowering their dividend payments by the corresponding dollar amount, hence keeping total payout constant at time 0. Then, the aggregate value of equities moves by a fraction*

$$v = \frac{\left(\mu^D - \mu^G\right)\theta}{\zeta + \mu^G\theta} b, \tag{56}$$

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

A provisional calibration Using the estimates of Di Maggio et al. (2020b), we set $\mu^D \simeq 0.5$ and $\mu^G \simeq 0.03$. Then, (56) says that a buyback of 1% of the market increases the market capitalization by 2.2%. The above papers (Baker et al. (2007); Di Maggio et al. (2020b)) do not exactly measure μ^D and μ^G: they measure the impact on consumption, not on consumption plus reallocation to pure bond funds. It is conceivable that some of the capital gains or dividends are reinvested in bonds, even if they're not consumed. So, μ^D (respectively μ^G) is likely to be higher than the marginal propensity to consume out of dividends (respectively capital gains). In addition, what matters is the "long run" propensity, which is hard to measure, and one may conjecture that the long-run consumption adjustment to a lasting policy change will have $\mu^D - \mu^G$ closer to 0. One upshot is that it would be interesting for the empirical literature to estimate the long-run μ^D and μ^G, as it is important to understand the impact of firms' actions such as buybacks in inelastic markets.

7 Conclusion

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

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

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

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

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

"Fast and smart investors (perhaps hedge funds) will provide enough elasticity to the market." This is not true: in part because hedge funds are small (they own less than 5% of the market, see Section 2), they cannot provide much elasticity for the market as a whole (so ζ remains low), even though they might ensure short term news are incorporated quickly (so that κ is quite high). In addition, those smart-money investors often face risk constraints and outflows that limit their ability to aggressively step in during aggregate downturns.

"Trading volume is very high, so the equity market must be very elastic." Trading volume in the equity market is high (about 100% of the value of the market each year), but most of it exchanges one share for another share (perhaps via a round-trip through cash). These trades within the universe of equities do no count toward the aggregate flow f, which is a (signed) flow from bonds to equities.

"For every buyer there is a seller; so, saying 'there was an increase in the demand for equities' is meaningless." Economists often appeal to the truism that "for every buyer there is a seller" to disregard the notion that a measurable increase in the willingness of the average trader to buy more of the market will push prices up ("buying pressure"). Our model clarifies that this reasoning is incorrect. In Proposition 2, f is the pressure to buy stocks (if it is positive), and the demand $q = -\zeta p + f$ has a component $-\zeta p$ expressing that "sellers" appetite to sell shares to "buyers" represented by f. So there are both buyers and sellers (or really, a force making the representative fund buy, and a force making it sell), but at the same time, buying pressure f does move the price by $p = \frac{f}{\zeta}$. Moreover, it is directly measurable via the change in asset holdings (bonds in the case of the undergraduate example of Section 3.1), as in (10).

"The market often looks impressively efficient in the short run, so it must be quite macro-efficient." The contrast between the market's "short run efficiency" and "macro-efficiency" is sharp in equation (20): future events are discounted at a rate $\rho = \frac{\zeta}{\kappa} = \delta + \frac{1-\theta}{\kappa}$, so that a highly far-sighted market has a lower value of ρ. So, the market can be very forward-looking (low ρ), even if it is very macro-inelastic (low ζ), provided that "far-sightedness" κ is relatively high compared to ζ (for example, because there are a few powerfully forward-looking arbitrageurs). As an example, consider the announcement of an event that will take effect in a week, such as a permanent increase

in dividends or inflows. In our calibration, the market's current reaction to the announcement is a fraction 99.8% of the eventual present value of the future dividends or inflows.[70] In that sense, the market looks impressively efficient. But again, it is "short-term predictability efficient" (it smooths announcements) and "micro efficient" (it processes well the relative valuations of stocks), but it is not "macro efficient" (as Samuelson (1998) put it) or "long-term predictability efficient" – it does not absorb well very persistent shocks. Furthermore, even though prices respond promptly around major events, it is generally hard to assess whether the market moved by just the right amount, or instead under- or over-reacted. In addition to a large literature demonstrating drifts in prices before and after macro events (such as Federal Open Market Committee meetings), our model implies that persistent flows around such events can lead to persistent deviations in prices, and typical event study graphs that do not display much of a drift in prices following the event would be uninformative about macro efficiency.

"Share buybacks do not affect equity returns, as proved by the Modigliani-Miller theorem." In the traditional frictionless model, the return impact of a share buyback should be zero. However, in our model, if firms in the aggregate buy back $1 worth of equity, that can increase aggregate valuations (Section 6.2 detailed this). Hence, share buybacks are potentially a source of fluctuations in the market. In our model, a combination of fund mandates and consumers' bounded rationality leads to a violation of the Modigliani-Miller neutrality. More broadly, corporate actions such as share issuances, transactions by insiders, et cetera, may have a large impact on prices beyond any informational channel. Most extant empirical evidence focuses on announcements at the firm level, while we emphasize their impact at the aggregate level. By focusing on well-identified firm-level responses, one identifies the micro-elasticity, not the macro elasticity ζ. It will be interesting to explore in detail how important corporate decisions are for fluctuations in the aggregate stock market.

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

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

Why is the aggregate demand for equities so inelastic? The core of the inelastic markets hypothesis is that the macro demand elasticity ζ is low. Why is it so low? We highlighted two reasons, namely fixed-share mandates (so that $\zeta > 0$, $\kappa = 0$), such as those of many funds that are 100% in equities and hence have zero elasticity, and inertia (i.e., some funds or people are just buy-and-hold, creating $\zeta = \kappa = 0$). This may be due to a taste for simplicity, or to agency frictions: as the household is not sure about the quality of the manager, a simple scheme like a constant share in equities may be sensible – otherwise the manager may take foolish risks.

There are other possibilities. If some funds have a Value-at-Risk constraint, and volatility goes up a lot in bad times, they need to sell when the markets fall, so that their ζ and κ are negative. A different possibility is that when prices move, people's subjective perception of the equity premium does not move much. One reason might be that investors think the rest of the market is well-informed. Also, going from market prices to the equity premium is a statistically

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

error-prone procedure, so that market participants may shrink towards no reaction to this (Black (1986), Summers (1986)). Alternatively, many investors may not place much weight on the price-earnings ratio as a reliable forecasting tool, perhaps because they want parsimonious models and price-dividend ratios are not that useful as short-run forecasters, or because many investors just do not wish to bother paying attention to them (Gabaix (2014), Chinco and Fos (2019)). The pass-through between subjective beliefs and actions might be low, as it is for retail investors (Giglio et al. (2021a)). Finally, demand may respond little to prices because demand shocks are highly persistent.[71] In the end, while identifying the exact reasons for low market elasticity is interesting, this question has a large number of plausible answers. Fortunately, it is possible to write a framework in a way that is relatively independent to the exact source of low elasticity, and this is the path we chose.

What are the determinants of flows? It is clear that it would be desirable to know more about the determinants of flows at a high frequency. We provided a minimalist model with a "behavioral disturbance" (which was enough to study its general equilibrium impact), and some simple correlations in Section 4.4, but this is clearly a first pass. Establishing the various channels of flows could be a whole line of enquiry, perhaps with micro data such as those used by Calvet et al. (2009) or Giglio et al. (2021a).

To appreciate the richness of those determinants, let us observe that flow shocks could come from various sources, such as: (i) changes in beliefs about future flows or fundamentals, as these both affect expected returns, per Proposition 5; (ii) "liquidity needs", for instance insurance companies selling stocks after a hurricane; (iii) more generally, heterogeneous income or wealth shocks to different groups (including foreign versus domestic investors) changing the effective propensity to invest in stocks by the average investor; (iv) corporate actions by firms such as decisions to buy back or issue shares; (v) shocks to substitute assets, which might for example prompt investors to rebalance towards stocks when bond yields go down; (vi) changes in the advertising or advice by institutional advisers, as explored in Ben-David et al. (2020b); (vii) "road shows" in which firms or governments try to convince potential investors to buy into a prospective equity offering or privatization; (viii) mechanical forced trading via "delta hedging," whereby traders who have sold put options and continuously hedge them need to sell stocks when stock prices fall.

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

[71] 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)).

and currencies): if so, how and what are the policy implications? This is a rich number of questions that hopefully economists will be able to answer in the coming years.

A Appendix: Main proofs

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

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

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

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

that is:

$$w_i = \theta_i p + f_i. \tag{57}$$

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

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

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

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

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

with $\zeta_i = 1 - \theta_i$. We see how $-\zeta_i$ is the (signed) demand elasticity, which includes crucial income effects.[72] For small price changes, this gives $q_i \simeq f_i - \zeta p_i$. We also see that, when $\kappa_i = 0$ for all funds, the equilibrium condition $q_S^D = 0$ leads to $p = \frac{f_S}{\zeta_S}$ exactly.

Next, consider the case with a general κ_i. Taking logs and then deviations from the baseline D/P ratio gives:

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

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

$$\hat{\pi} = \delta\,(d-p)\,. \tag{58}$$

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

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

$$q_i = w_i - p + \kappa_i \hat{\pi}. \tag{59}$$

Using (57), we obtain (7):

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

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

$$w_t = \theta p_t + f_t. \tag{60}$$

Now, from the demand for stocks, we have $Q_t P_t = W_t \theta e^{\kappa \hat{\pi}_t + \nu_t}$, while in the baseline economy $\bar{Q}_t \bar{P}_t = \bar{W}_t \theta$. Dividing through, we get: $\frac{Q_t P_t}{\bar{Q}_t \bar{P}_t} = \frac{W_t}{\bar{W}_t} e^{\kappa \hat{\pi}_t + \nu_t}$, so that $(1 + q_t)(1 + p_t) = (1 + w_t)\, e^{\kappa \hat{\pi}_t + \nu_t}$. Linearizing, $q_t + p_t = w_t + \kappa \hat{\pi}_t + \nu_t$. Hence, by (60),

$$q_t = -(1 - \theta)\, p_t + \kappa \hat{\pi}_t + f_t + \nu_t. \tag{61}$$

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

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

$$\mathbb{E}_t \Delta p_{t+1} - \rho p_t + \delta d_t^e + \frac{f_t + \nu_t}{\kappa} = 0. \tag{62}$$

Defining $z_t := \delta d_t^e + \frac{f_t + \nu_t}{\kappa}$, this gives $p_t = \frac{\mathbb{E}_t p_{t+1} + z_t}{1 + \rho}$, so that $p_t = \mathbb{E}_t \sum_{\tau \geq t}^{\infty} \frac{z_\tau}{(1+\rho)^{\tau - t + 1}}$. The equity premium comes from (61) with $q_t = 0$.

B Appendix: Identification methodology

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

B.1 Algorithm used for sector-level data

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

1. We construct pseudo-equal value weights $\tilde{E}_{i,t-1}$, where we start from $\tilde{E}_i^{\sigma} = \frac{\sigma_i^{-2}}{\sum_{k=1}^{N} \sigma_k^{-2}}$, where $\sigma_i = \sigma(\Delta q_{it})$, and define $\tilde{E}_i = \min\left\{\xi \tilde{E}_i^{\sigma}, \frac{1.5}{N}\right\}$, where $\xi \geq 1$ is tuned so that $\sum_i \tilde{E}_i = 1$. We exclude the corporate sector in constructing the instrument. This winsorizes the quasi-equal weights to be at most 50% higher than strict equal weights. This adjustment ensures that the equal weights are not too concentrated for sectors with very stable Δq_{it}.[73] This is relevant when the number of sectors is small, as is the case for the FoF.

2. We run the panel regression

$$\Delta q_{it} = \alpha_i + \beta_t + \gamma_i \Delta y_t + \delta_i t + \Delta \check{q}_{it}, \tag{63}$$

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

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

4. We construct the GIV instrument:[74]

$$Z_t = \sum_{i=1}^{N} S_{i,t-1} \Delta \check{q}_{it}. \tag{64}$$

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

$$\Delta p_t = \alpha + M Z_t + \lambda_P' \eta_t^e + e_t, \tag{65}$$

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

$$\Delta q_{Et} = \alpha_E - \zeta \Delta p_t + \lambda_E' \eta_t^e + e_t, \tag{66}$$

and the supply elasticity via

$$\Delta q_{Ct} = \alpha_C - \zeta_C \Delta p_t + \lambda_C' \eta_t^e + e_t. \tag{67}$$

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

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

B.2 Algorithm used for investor-level data

We summarize the algorithm that we use to extract factors, η_t, in Section 4.3.

1. We run the panel regression

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

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

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

3. We extract principal components based on the panel of 20 groups of $\Delta \check{q}^E_{gt}$.

References

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

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

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

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

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

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

Barberis, Nicholas, Ming Huang, and Richard H Thaler, "Individual preferences, monetary gambles, and stock market participation: A case for narrow framing," *American economic review*, 2006, *96* (4), 1069–1090.

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

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

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

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, "Do ETFs Increase Volatility?," *The Journal of Finance*, 2018, *73* (6), 2471–2535.

Ben-David, Itzhak, Francesco Franzoni, Rabih Moussawi, and John Sedunov, "The granular nature of large institutional investors," *Management Science*, forthcoming.

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

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

Ben-Rephael, Azi, Shmuel Kandel, and Avi Wohl, "Measuring investor sentiment with mutual fund flows," *Journal of Financial Economics*, 2012, *104* (2), 363–382. Special Issue on Investor Sentiment.

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

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

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

Bouchaud, Jean-Philippe, Julius Bonart, Jonathan Donier, and Martin Gould, *Trades, quotes and prices: financial markets under the microscope*, Cambridge University Press, 2018.

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

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

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

Buffa, Andrea M, Dimitri Vayanos, and Paul Woolley, "Asset management contracts and equilibrium prices," Technical Report, National Bureau of Economic Research 2019.

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

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

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

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

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

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

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

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

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

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

Cole, Allison, Jonathan Parker, and Antoinette Schoar, "Household Portfolios and Retirement Saving Over the Life Cycle," 2021. Working Paper, MIT Sloan.

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

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

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

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

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

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

Di Maggio, Marco, Amir Kermani, and Kaveh Majlesi, "Stock Market Returns and Consumption," *Journal of Finance*, 2020, *75* (6), 3175–3219.

Dierker, Martin, Jung-Wook Kim, Jason Lee, and Randall Morck, "Investors' Interacting Demand and Supply Curves for Common Stocks," *Review of Finance*, 2016, *20* (4), 1517–1547.

Dong, Xi, Namho Kang, and Joel Peress, "How the Speed of Capital Affects Factor Momentum, Reversal and Volatility?," *Available at SSRN*, 2021.

Dou, Winston, Leonid Kogan, and Wei Wu, "Common Fund Flows: Flow Hedging and Factor Pricing," *Available at SSRN*, 2020.

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

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

Edelen, Roger M and Jerold B Warner, "Aggregate price effects of institutional trading: a study of mutual fund flow and market returns," *Journal of Financial Economics*, 2001, *59* (2), 195–220.

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

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

Frazzini, Andrea, Ronen Israel, and Tobias J Moskowitz, "Trading costs," *Working Paper*, 2018.

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

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

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

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

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

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

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

Gabaix, Xavier and Ralph SJ Koijen, "Granular instrumental variables," Working Paper 28204, National Bureau of Economic Research December 2020.

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

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

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

Ghysels, Eric, Hanwei Liu, and Steve Raymond, "Institutional Investors and Granularity in Equity Markets," 2021. Working Paper.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, "Five Facts About Beliefs and Portfolios," *American Economic Review*, 2021, *111* (5), 1481–1522.

Goetzmann, William N. and Massimo Massa, "Index Funds and Stock Market Growth," *The Journal of Business*, 2003, *76* (1), 1–28.

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

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

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

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

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

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

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

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

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

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

Kahneman, Daniel and Dan Lovallo, "Timid choices and bold forecasts: A cognitive perspective on risk taking," *Management science*, 1993, *39* (1), 17–31.

Kekre, Rohan and Moritz Lenel, "Monetary policy, redistribution, and risk premia," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2020, (2020-02).

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

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

Koijen, Ralph SJ and Motohiro Yogo, "Exchange Rates and Asset Prices in a Global Demand System," 2020. NBER Working Paper 27342.

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

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

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

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

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

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

Li, Jiacui, "What Drives the Size and Value Factors?," 2021. Working Paper, University of Utah.

Lou, Dong, "A Flow-Based Explanation for Return Predictability," *The Review of Financial Studies*, 12 2012, *25* (12), 3457–3489.

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

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

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

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

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

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

Parker, Jonathan A, Antoinette Schoar, and Yang Sun, "Retail Financial Innovation and Stock Market Dynamics: The Case of Target Date Funds," *Working Paper*, 2020.

Pavlova, Anna and Taisiya Sikorskaya, "Benchmarking Intensity," *Working Paper*, 2020.

Peng, Cameron and Chen Wang, "Factor demand and factor returns," *Working Paper*, 2021.

Petajisto, Antti, "Why do demand curves for stocks slope down?," *Journal of Financial and Quantitative Analysis*, 2009, *44* (5), 1013–1044.

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

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

Schmickler, Simon, "Identifying the Price Impact of Fire Sales Using High-Frequency Surprise Mutual Fund Flows," *Working Paper*, 2020.

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

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

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

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

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

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

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

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

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

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