

# Stock Demand and Price Impact of 401(k) Plans\*

Riccardo Sabbatucci<sup>†</sup>

Stockholm School of Economics & SHoF

Andrea Tamoni<sup>‡</sup>

Rutgers Business School

Song Xiao<sup>§</sup>

LSE

March 14, 2023

## ABSTRACT

We estimate a demand system linking 401(k) plans ownership of individual stocks and funds to their demand for equities, and quantify the effect of 401(k) assets on fund managers' investment behavior. We find that 401(k) fund and stock ownership are the most important variables, after size, explaining fund demand for stocks, with a one standard deviation increase in 401(k) ownership leading to 15-30% increase in stock demand. Funds managing a larger fraction of 401(k) assets tilt their portfolios toward winners, high beta and long duration stocks, outperforming their benchmarks. This investment behavior has important implications for security pricing and generate a feedback effect if pension flows respond positively to relative fund returns. Lastly, we estimate the equilibrium price impact of a change in 401(k) ownership to be positive and increasing over time, consistent with the shift from active to passive investing.

**Keywords:** 401(k) plans; stock demand and 401(k) ownership; demand based asset pricing; price impact of 401(k) plans

---

\*We are grateful to Ralph Koijen, David Solomon for valuable suggestions on our paper. We also received helpful comments from seminar participants at University of Miami, Herbert Business School.

<sup>†</sup>Department of Finance, Stockholm School of Economics and Swedish House of Finance (SHoF). **Email:** [riccardo.sabbatucci@hhs.se](mailto:riccardo.sabbatucci@hhs.se)

<sup>‡</sup>Department of Finance, Rutgers Business School. **Email:** [atamoni@business.rutgers.edu](mailto:atamoni@business.rutgers.edu)

<sup>§</sup>PhD Student, Department of Finance, London School of Economics. **Email:** [s.xiao2@lse.ac.uk](mailto:s.xiao2@lse.ac.uk)

# 1 Introduction

What drives fluctuations in the valuation of different asset classes? A recent and influential strand of the literature has tried to answer this question using a demand-based framework, following the seminal work of [Kojien and Yogo \(2019\)](#). Demand-based explanations linked to the role of heterogeneity in investor types have been recently proposed to explain variation in domestic equities ([Kojien and Yogo, 2019](#)), bonds ([Bretscher et al., 2021](#)), exchange rates ([Kojien and Yogo, 2020](#)), and to analyze the shift from active to passive and green investments ([Kojien et al., 2022](#)), and the portfolios of high-net worth individuals ([Gabaix, Kojien, Mainardi, Oh and Yogo, 2022](#)).

One of the key elements of the demand-based asset pricing framework is the role of the latent demand, defined as the investor-specific demand not explicitly captured by well-known stock and investor characteristics.

In this paper, we shed some light on this important component for a specific class of institutional investors, active fund managers. More precisely, using a demand-based framework, we study the impact that 401(k) pension plans have on fund demand for stocks. We introduce two plausible channels through which 401(k) ownership of fund assets and stocks might affect investors' allocation decisions.

The first channel through which 401(k) allocations affect stock demand is by direct flows to non-index tracking mutual funds and ETFs, which, in turn, use that additional liquidity to increment their equity exposure. In other words, funds with the largest fraction of 401(k) assets might have more stable flows, and hence invest in different types of stocks compared to funds managing fewer 401(k) assets. For example, funds might display preference for stocks with high market beta (e.g., [Christoffersen and Simutin \(2017\)](#), [Han et al. \(2022\)](#)). We label this the *fund level* channel.

The second channel through which 401(k) plans affect the demand of individual stocks is by ownership of the stock itself. The fraction of an individual stock owned by 401(k) plans is an additional stock characteristic, similarly to book-to-market or momentum. Fund managers might take into account the information embedded into this additional stock characteristic when evaluating how many shares of a specific stock to purchase.

Hence, the ownership of a stock by 401(k) plans might affect the fund demand for that specific stock. We label this the *stock level* channel.

We report four main findings. First, funds managing a larger fraction of 401(k) assets display larger demand for stocks. More in detail, 401(k) fund ownership is the most important variable, after firm's equity, in explaining how much of a stock funds demand. For a standard deviation increase in 401(k) fund ownership, the average mutual fund demands approximately 33.3% (*t*-stat: 2.91) more of the average stock, while the demand from ETFs is almost unchanged. This test of the fund channel mechanism suggests that mutual fund managers take into account the amount of 401(k) assets they manage when deciding their portfolio allocation, and have more discretion than ETFs managers.

Second, we test the stock-level channel mechanism, and find that the amount of company shares owned by 401(k) plans is an important characteristic – in fact, the most important one after firm's equity – in explaining the decision of mutual funds and ETFs to invest in a stock. For a one standard deviation increase in 401(k) stock ownership, the average mutual fund demands approximately 18.6% (*t*-stat: 11.16) more of the stock. Differently from the fund-level case, the average ETF also increases exposure to the stock by 11.5% for each standard deviation increase in 401(k) stock ownership. A natural question is why ETFs choose stocks with large 401(k) ownership. On the one hand, it could be that the stock-level 401(k) characteristic is correlated with the probability of the stock being included in the benchmarks tracked by ETFs (e.g., S&P500), since stocks with large institutional 401(k) ownership have already been screened and vetted by investors that bought them, and hence are included in most benchmarks. On the other hand, it might be that some ETFs are not perfectly tracking their respective benchmarks, and hence have some leeway in choosing their stock holdings. Indeed, we find this to be the case, as 30% of the ETFs in our sample display a tracking error of more than 2.18% per year with respect to their benchmarks, suggesting that many ETF managers still have flexibility in selecting their stock holdings. These two sets of results suggest that 401(k) flows are important to fund managers, and shape their investment decisions.

Third, we study how the investment strategies and performance of mutual funds are affected by the amount of 401(k) assets they manage. First, we analyze the 401(k) asset-

induced fund demand for specific stock characteristics. We find that fund demand for stocks is heterogeneous, as a function of 401(k) fund-level ownership. Funds managing a larger fraction of 401(k) assets tilt their portfolios toward winners, high beta and long duration stocks, and away from large stocks. This fund behavior can reinforce the well known betting-against-beta ([Frazzini and Pedersen, 2014](#)) and duration anomalies. Second, we study the relative and risk-adjusted performance of a fund as a function of 401(k) fund-level ownership. We find that a large fraction of pension assets managed by funds improves their performance in terms of relative returns, but leaves (statistically) unaffected the alpha. This result is important: if pension plans choose relative returns as the main criterion for investment (rather than alpha), the better relative performance of funds with high 401(k) ownership can induce positive pension flows, triggering a spiral effect. Furthermore, the preference of fund managers for high beta and long duration stocks (holding alpha constant) provides support for the literature on benchmarking and manager incentives ([Baker et al., 2011](#); [Buffa et al., 2022](#)).

Lastly, we estimate the equilibrium price impact of a change in the stock-level 401(k) characteristic for the cross-section of stocks, and over time. We find the institutional price pressure to be positive and increasing over our sample. For the median stock, the price impact raises from 0.2 in 2007 to 0.6 in 2020. For stocks lying on the 90th percentile of the price impact distribution, it hovers around 0.8 over our sample. We also compute the price impact for portfolios of stocks sorted on size, book-to-market, or beta (market risk). We find that the average price impact as a function of stock-level 401(k) has increased for large stocks, while it has remained relatively stable for small stocks. However, we do not observe noticeable differences for stocks sorted on book-to-market or betas, i.e., they are equally impacted. The positive trend of price impact, which increases almost monotonically between 2008 and 2020 is consistent with the shift from active to passive investing of the last decade documented in the literature ([Kojien et al., 2022](#)).

Our paper is related to the emerging demand-based asset pricing literature. [Kojien and Yogo \(2019\)](#) develop the demand system approach and document that changes in latent demand (e.g., characteristics unobserved by the econometrician) are explaining 81 percent of the cross-sectional variance of stock returns. [Bretscher et al. \(2021\)](#) and [Gabaix](#)

et al. (2022) estimate a demand system for corporate bonds and for high-net-worth investors, respectively. Koijen et al. (2022) use the demand-based system to study the impact of market trends, such as the shift from active to passive investing or the increased demand for green firms, on price informativeness. Haddad et al. (2022) investigate the effect of the switch to passive investing, and document that this behavior has led to substantially more inelastic aggregate demand curves for individual stocks.

Our contribution to this strand of the literature is to highlight the relevance of 401(k) plans ownership – both at the fund and individual stock levels – in driving fund managers’ investment decisions.

Our paper is also related to the literature on risk preferences and shifting of fund managers. Christoffersen and Simutin (2017) show that funds controlling large pension assets tend to increase their exposure to high beta stocks. Differently from this paper, we estimate the stock demand of funds as a function of 401(k) plan ownership, controlling for stock characteristics.<sup>1</sup> Han et al. (2022) document that underperforming funds increase their demand for risky stocks. Our contribution relative to this study is to emphasize the role of 401(k) defined contribution pension assets in determining fund managers’ risk profile, e.g., the *types* of stocks demanded by fund managers. Dou et al. (2022) show that active funds care about their size, which is affected by fund flows that obey a strong factor structure with the common component responding to macroeconomic shocks. They find that high-flow-beta stocks earn significantly higher excess returns and higher capital asset pricing model (CAPM) alphas in the cross-section. Relative to their work, we document that an important component of fund size is determined by 401(k) plans allocation, and that 401(k) ownership affects the types of stocks preferred by fund managers.

Lastly, our paper is also related to the literature on pension plans. Sialm and Starks (2012) study the investment strategies and performance of funds held primarily by retirement accounts versus those held by taxable investors. They do not find performance differences between funds held by different tax clienteles. Differently from their paper, we document that a large fraction of pension assets managed by a fund influences its per-

---

<sup>1</sup>Christoffersen and Simutin (2017), uses survey data from “Pensions & Investments” (P&I) administered to domestic equity funds for the years 2004 through 2014, while we use actual data on annual 401(k) plan holdings.

formance and asset selection. [Sialm et al. \(2015\)](#) study the investment menu of pension plans, and find that flows into funds from defined contribution (DC) assets are less sticky and more sensitive to fund performance than non-DC flows, because of adjustments to the investment options by the plan sponsors. They document that plan participants exhibit inertia and do not react sensitively to prior fund performance. Differently from their work, we do not focus on investment menu offered by plans, but directly estimate the demand of individual stocks by funds available in 401(k) plans using a demand-based framework.

Differently from [Sialm et al. \(2015\)](#) and [Christoffersen and Simutin \(2017\)](#), we do not rely on surveys about defined contribution assets, but observe the actual 401(k) plan holdings using a novel dataset, Brightscope. Using the same data, [Egan et al. \(2021\)](#) address a different research question, and document heterogeneity in investment behavior of 401(k) participants, showing that higher income and more educated individuals tend to have higher equity exposure, whereas retirees and minorities tend to have lower equity exposure.

The remainder of the paper proceeds as follows. [Section 2](#) describes the institutional framework and the data used in the paper, while [Section 3](#) introduces our demand-based framework and our testable hypotheses. [Section 4](#) presents the empirical analysis and discusses our findings. [Section 5](#) concludes.

## 2 Data

### 2.1 Data Sources

Our detailed 401(k) plan holding data comes from BrightScope Beacon. BrightScope Beacon provides comprehensive plan-level holding data gathered from audited Form 5500 filings of private-sector defined contribution (DC) plans from 2006 to 2020. We focus exclusively on 401(k) defined contribution plans in this paper.<sup>2</sup> BrightScope reports annual data on the investment options (e.g., mutual funds) available to plan participants together

---

<sup>2</sup>BrightScope Beacon also provides holdings for 403(b) plans, although their total market value is small relative to that of 401(k) plans.

with the total dollar amount invested in each option. In other words, for each 401(k) plan, we observe its asset allocation on equity mutual funds (including ETFs), allocation funds (including TDFs), bond mutual funds and other types of assets (e.g., trusts and common stocks), over time. The dataset covers 708,929 different 401(k) plans over the period 2006-2020, resulting in more than 8 million fund-by-plan-by-year observations. In addition, data on fund names, fees, and tickers is also available.

Mutual fund holdings and characteristics, such as their expense ratio, category, fund domicile, investment type (e.g., ETF flag), AUM, and tickers, are obtained from Morningstar Direct.<sup>3</sup> We match mutual funds in 401(k) plans with Morningstar by fund tickers and names.<sup>4</sup>

Given our interest in the impact of 401(k) plans on US stocks, we focus on domestic equity mutual funds. Specifically, we keep mutual funds with equity ratios greater than 0.75 and remove non-US equity funds based on the Morningstar fund domicile variable.<sup>5</sup> We also require funds to have at least 3 years of holdings data. We only include equity mutual funds and ETFs directly owned by 401(k) plans.<sup>6</sup> Our final dataset is thus comprised of a total of 2,156 funds, 1,763 mutual funds and 393 ETFs.

Lastly, we supplement the Morningstar holdings data with stock data from CRSP and Compustat. In our empirical analysis, we use the same stock characteristic as in [Kojien and Yogo \(2019\)](#), namely, log book equity, profitability, investment, dividends-to-book equity and market beta, in addition to the instrumented log market-to-book ratio, as in [Kojien et al. \(2022\)](#). Profitability is defined as operating profits scaled by book value of equity, investment as the annual growth rate of total assets, and dividends-to-book equity

---

<sup>3</sup>Morningstar provides exhaustive mutual fund holdings compared to other mutual fund holding databases, such as CRSP. [Schwarz and Potter \(2016\)](#) find that CRSP misses many SEC mandated portfolios available in SEC filings.

<sup>4</sup>More precisely, we map mutual fund tickers in BrightScope Beacon to Morningstar mutual fund ID (variable: *fundid*) when tickers are available in both datasets. When fund tickers are missing in either dataset, we match mutual funds by their names. We match 98.2% of mutual fund allocation in retirement plans, or a total of 3,182 mutual funds and ETFs.

<sup>5</sup>Additionally, we remove mutual funds whose portfolio weights reported by Morningstar are different from the correct portfolio weights calculated using holdings values and total net assets, as in [Pástor et al. \(2015\)](#).

<sup>6</sup>Target-date funds also invest in mutual funds and ETFs, but their rebalancing between equity and bonds is mechanical as a function of fund age. Hence, we only focus on funds directly selected by 401(k) pension plans.

as the ratio of annual dividends to book equity. Stock market beta is estimated using a 60-month rolling window regression of monthly stock excess returns, over the 1-month Treasury-bill rate, on market excess returns, with at least 20 months of non-missing observations. Fund TNAs are winsorized at the 99th percentile at the end of every year to limit the impact of outliers. [Internet Appendix C](#) describes the data cleaning procedures in detail.

## 2.2 Descriptive Statistics

[Figure 1](#) displays the allocation of 401(k) plans to the various investment categories. These include direct ownership of individual stocks, separate accounts, guaranteed investment contracts (GIC),<sup>7</sup> mutual funds (including ETFs) and collective investment trusts (CIT).

Collective investment trusts (CIT), the second largest component, averaging 24% of 401(k) assets under management over our sample period, are pooled investment vehicles established by banks or trust companies, that are only available to defined-contribution (DC) plan participants when the CITs are included as options in the DC plan menu.<sup>8</sup> The Goldman Sachs Core Plus Fixed Income (bonds) and T. Rowe Price Blue Chip Growth Trust (equity) are two examples of CIT options offered by large investment companies to DC plan sponsors.

The mutual fund category, which also includes ETFs, is the largest component with, on average, 43% of the total 401(k) assets. [Figure 2](#) decomposes the category into five groups: US equity ETFs, US equity mutual funds, US index funds, allocation funds, and others. Allocation funds include target-date funds and balanced funds investing in a mix of equity and fixed income assets, while international mutual funds, bond mutual funds, money market mutual funds, and alternative investment funds are pooled together in

---

<sup>7</sup>GICs are agreements between an investor and an insurance company, typically available in retirement plans, whereby the insurance company guarantees the investor a certain rate of return in exchange for holding the deposit for a fixed period of time.

<sup>8</sup>Differently from mutual funds, CITs are not required to publicly disclose holdings. Moreover, while mutual funds can be bought by most investors through, for example, a brokerage firm, 100% of CIT assets linked to a DC plan can only be owned by DC participants. Therefore, even if the holdings were available, we would not be able to estimate the marginal impact of 401(k) ownership on CITs demand for stocks since there is no cross-sectional variation in 401(k) ownership across CITs.



the “Others” category. US index funds include mutual funds and ETFs that are index-tracking.<sup>9</sup>

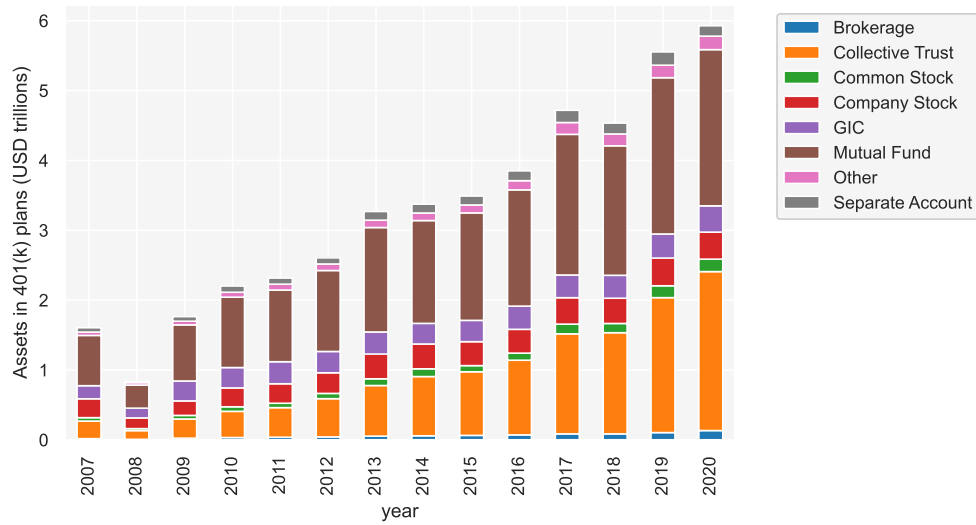
Our focus is on the two remaining groups, active mutual funds and ETFs investing in US equities. We observe a substantial increase in mutual funds (orange bar) and ETFs (green bar) assets over time, with the former (latter) totaling around \$0.63tn (\$32bn) in 401(k) as of 2020. Active ETFs assets inside 401(k) plans are still somehow limited, but they are growing at the fastest rate over the last 5 years. In fact, the annual growth of ETF investments by 401(k) plans amounts to 16% over the last five years, and 25% over the period 2007-2020.

[Table 1](#) reports the cross-sectional distribution, across years, of some 401(k) plan characteristics. The first variable,  $IO^{401k}$ , indicates the fraction of assets of an individual fund owned collectively by 401(k) plans. We observe that around 8% of the fund assets are owned, on average, by 401(k) plans, making 401(k) plans among the largest fund investors. Most importantly, the dollar amount invested by a given 401(k) plan in a specific fund, as a fraction of the total plan assets, is quite persistent. When looking at the top 25% of the plan-fund distribution, we observe an annual autoregressive coefficient of 0.82. The third row displays the size distribution across 401(k) plans. We find that the average 401(k) plan size is around \$92mn, while the median is only \$11mn, suggesting that the cross-sectional distribution is extremely right skewed, consistent with [Egan et al. \(2021\)](#). The last three rows report 401(k) plans’ dollar allocation to index funds, active US equity mutual funds, and ETFs, respectively, and show that 401(k) plans invest substantially more in mutual funds than ETFs over our sample period, while allocating a relevant fraction of their assets to equity index funds.

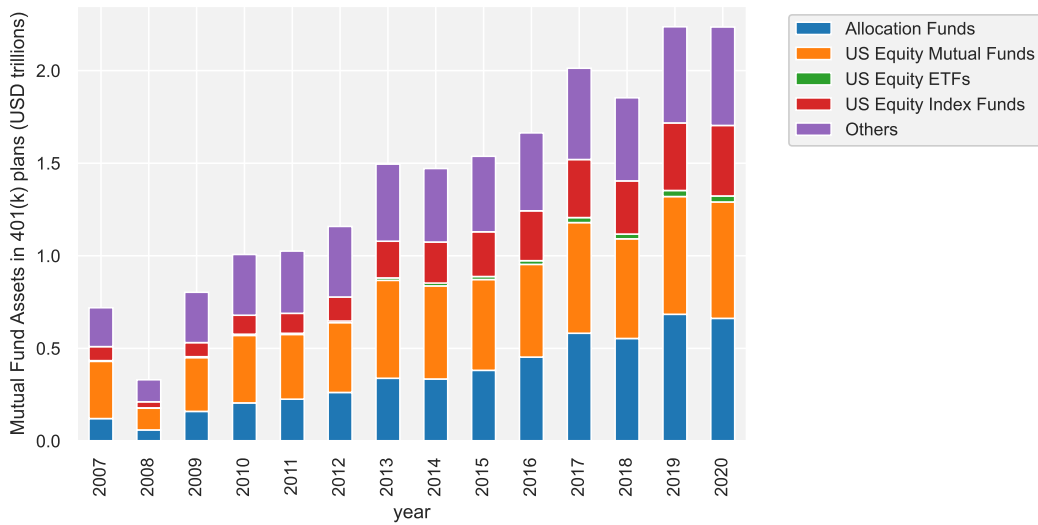
[Figure 3](#) shows the cross-sectional distribution of the fund-level 401(k) ownership,  $IO_{i,t}^{401k}$ , over time. We notice that the ownership is stationary, even at the 75th percentile of the distribution where it fluctuates between 6% and 12%.

---

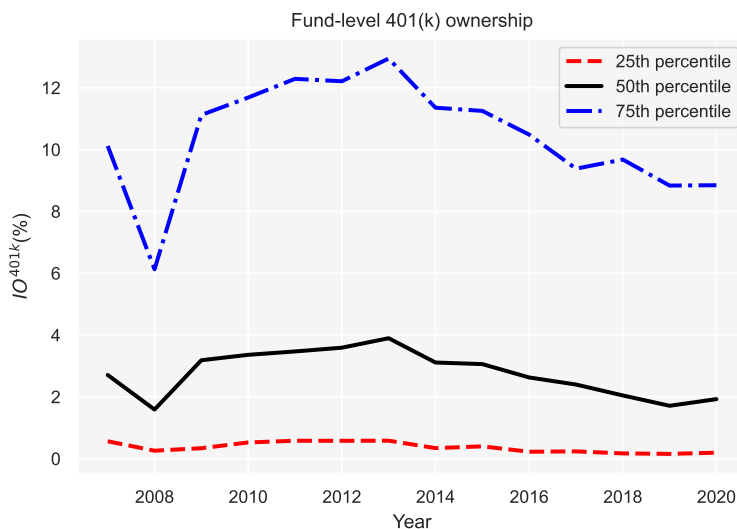
<sup>9</sup>Specifically, we define index-tracking ETFs as large cap ETFs that track the S&P500 index (based on Lipper code: “SP”, S&P 500 Index Objective Funds). We define index mutual funds according to the Morningstar classification (e.g., index funds and enhanced index categories).



**Figure 1: 401(k) plan assets.** This figure shows the distribution of 401(k) plan assets into the various investment options, over time. Annual data, from 2007 to 2020.



**Figure 2: Distribution of assets within the mutual fund category.** This figure plots the value of 401(k) mutual fund investments split into various subgroups. Allocation funds are balanced funds investing in a mix of fixed income assets and equities depending on their objective, e.g., target-date funds. US equity mutual funds and ETFs include all active domestic equity funds. US equity index funds include both mutual funds and ETFs that are index-tracking. The category “Others” includes bond mutual funds, international equity mutual funds, money market funds and alternative investment funds.



**Figure 3: Fund-level 401(k) ownership over time.** This figure shows the cross-sectional distribution of fund-level 401(k) ownership over time. Annual data, from 2007 to 2020.

### 3 Estimating the Impact of 401(k) Plans on Stock Demand

As discussed in [Section 2](#), 401(k) plans hold a substantial amount of assets invested in equity mutual funds and ETFs. In this section, we outline how we adapt the asset demand framework of [Kojien and Yogo \(2019\)](#) for our purpose, and highlight the two main channels through which retirement plan allocations can impact the demand of mutual funds and ETFs for individual stocks.

#### 3.1 Model

We extend [Kojien and Yogo \(2019\)](#), and define the demand curve of investor  $i$  for stock  $n$ <sup>10</sup> as:

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = \exp \left\{ b_{0,i,t} + \beta_{0,i} mb_t(n) + \beta'_{1,i} \mathbf{X}_t(n) + \beta_{2,i} IO_{i,t}^{401k}(n) \right\} \epsilon_{i,t}(n) \quad (1)$$

where  $mb_t(n)$  is the log market-to-book equity of asset  $n$  at time  $t$ ,  $\mathbf{X}_t(n)$  is a vector of

<sup>10</sup>In our setting, we consider two groups of investors: US equity mutual funds and ETFs.

$k$  observed characteristics of asset  $n$  at date  $t$ , and  $w_{i,t}(0)$  is the portfolio weight on the outside asset. As in [Kojien and Yogo \(2019\)](#), we include log book equity, profitability, investment, dividend-to-book equity, and market beta as characteristics. In addition to the aforementioned stock characteristics, we augment the original model in [Kojien and Yogo \(2019\)](#) with one additional variable, potentially capturing variation in investor demand: 401(k) ownership of either the investor  $i$  (denoted  $IO_{i,t}^{401k}$ ), or the individual stock  $n$  (denoted  $IO_t^{401k}(n)$ ). In our setting, the underlying hypothesis is that the fraction of mutual fund/ETF assets owned by defined contribution (401(k) plans) is a key determinant of fund managers' stock demand. Empirically, the relative importance of 401(k) assets in driving the latter will be the studied in [Section 4](#).

Following [Kojien and Yogo \(2019\)](#), we assume throughout that stock characteristics are exogenous to latent demand,

$$\mathbb{E}_t \left[ \epsilon_{i,t}(n) \mid \mathbf{X}_t(n), IO_{i,t}^{401k}(n) \right] = 1. \quad (2)$$

By explicitly controlling for variables such as book-to-market and profitability in the regressions, and instrumenting market equity as in [Kojien and Yogo \(2019\)](#) (see next paragraph), we limit potential endogeneity concerns for the fund-level 401(k) ownership. In other words, stock-specific characteristics that affect the behavior of fund managers, other than those explicitly used as regressors in the model, are unlikely to affect the allocation of 401(k) plans to funds, and their choice of individual stocks. However, 401(k) plans may select funds based on fund-specific characteristics, such as the fund style (e.g., "growth"), its manager, and the size of the fund. [Section A.2](#) provides robustness tests for the exogeneity of the fund-level 401(k) ownership along these additional dimensions.

However, latent investor demand is likely correlated with a stock's market capitalization, i.e.,  $\mathbb{E}_t [\epsilon_{i,t}(n) \mid me_t(n)] \neq 0$ , because some investors are large and their individual latent demand impacts stock prices.<sup>11</sup> Hence, the model in (1) delivers biased and inconsistent estimates.

We therefore construct an instrument  $z_{i,t}(n)$  for the endogenous market capitalization.

---

<sup>11</sup>Market equity is the numerator of log market-to-book equity.

Specifically, we follow [Kojien et al. \(2022\)](#) and use exogenous variation in investors' investment mandates to generate exogenous variation in demand. Let  $\mathcal{S}_{i,t}$  denote the set of stocks held in period  $t$  and assume that any stock that investor  $i$  holds during the current year, or any of the previous 11 quarters, is part of her choice set,  $\mathcal{N}_{i,t} = \cup_{k=0}^{11} \mathcal{S}_{i,t-k}$  where  $k$  is expressed in years.<sup>12</sup>

Note that if  $n \notin \mathcal{N}_{i,t}$ , it means that stock  $n$  is part of the outside asset for investor  $i$  at time  $t$ . When  $w_{i,t}(n) = 0$ , stock  $n$  belongs to the investment universe of investor  $i$ , but she does not hold the stock at time  $t$ , hence the characteristics-based demand in (1) is able to account for zero holdings of a stock. In order to construct an instrument that relies only on the fund investment universe, but not on the exact investor  $i$  holdings within the investment universe, we compute counterfactual market equity  $z_{i,t}(n)$  (i.e., the instrument) as if investors held an equal-weighted portfolio of all the stocks in their investment universe, and excluding the investor's own holdings.<sup>13</sup>

$$z_{i,t}(n) = \log \left( \sum_{j \neq i} A_{j,t} \frac{1_{n \in \mathcal{N}_{j,t}}}{1 + |\mathcal{N}_{j,t}|} \right)$$

where  $1_{n \in \mathcal{N}_{j,t}}$  is an indicator function equal to one if the stock  $n$  belongs to investor  $j$ 's choice set  $\mathcal{N}_{j,t}$ ,  $A_{i,t}$  denotes the dollar assets owned by investor  $i$  at time  $t$ , and  $|\mathcal{N}_{j,t}|$  denotes number of stocks in an investor's choice set.<sup>14</sup>

## 3.2 Economic Channels

The first channel through which 401(k) allocations affect stock demand is by direct flows to mutual funds and ETFs, which, in turn, use that additional liquidity to increment their equity exposure. Since 401(k) plans tend to be low turnover investors, especially relative to other types of institutional investors, funds with the largest 401(k) ownership might

<sup>12</sup>For each fund  $i$ , the outside asset includes the complement set of stocks, those not in the investment universe.

<sup>13</sup>Since we focus on US equity mutual funds and ETFs, we use their investment universe. Specifically, the summation in  $z_{i,t}(n)$  spans all the mutual funds and ETFs that are held by retirement plans.

<sup>14</sup>Although there are  $|\mathcal{N}_{j,t}| + 1$  assets including the outside asset, there are only  $|\mathcal{N}_{j,t}|$  degrees of freedom implied by the budget constraint, since asset weights must sum to unity.

display a more stable investor base.<sup>15</sup> As a consequence, funds managing a larger fraction of 401(k) assets might invest in different types of stocks compared to funds with fewer 401(k) assets, all else being equal. In other words, the “investor base” of mutual funds/ETFs might affect funds’ asset allocation decisions, e.g., funds may increase exposures to specific stock characteristic (Christoffersen and Simutin, 2017), opting perhaps for riskier bets.<sup>16</sup> We label this the *fund level* channel.

The second channel through which 401(k) plans affect the demand of individual stocks is by ownership of the stock itself.<sup>17</sup> We can think of the percentage of an individual stock held by 401(k) plans as an additional stock characteristic, similarly to book-to-market or momentum. Fund managers might take into account the information embedded into this additional stock characteristic when evaluating how many shares of a specific stock to purchase.<sup>18</sup> Hence, the total percentage ownership of a stock by 401(k) plans might affect the fund demand for that specific stock. We label this the *stock level* channel.

Koijen and Yogo (2019) highlight the importance of latent asset demand, defined as the component of the demand function unexplained by the model covariates. We conjecture that an important component of the variation in this latent asset demand is attributable to the two economic channels described above, e.g., the fraction of fund  $i$ ’s assets under management owned by all 401(k) plans at time  $t$ , and the fraction of stock  $n$  owned in aggregate by 401(k) plans. We next estimate the magnitude of these effects.

## 4 Empirical Results

### 4.1 Fund Level Channel

We define the fraction of fund  $i$ ’s assets under management owned by aggregate 401(k) plans at time  $t$  as

---

<sup>15</sup>Recall from Section 2 that the top quartile of 401(k) plans have allocations to funds that are persistent in percentage terms, with an annual autoregressive coefficient larger than 0.8.

<sup>16</sup>In Section 4.3 we test the risk preferences of fund managers as a function of the 401(k) assets they control.

<sup>17</sup>We only focus on indirect ownership, e.g., through funds, since direct stock ownership by 401(k) plans is usually negligible.

<sup>18</sup>Institutional ownership of a stock is a well known characteristic amongst investors, and 401(k) plans ownership can be retrieved by public filings or third party data providers.

$$IO_{i,t}^{401k} = \frac{\sum_{p=1}^M AUM_{p,i,t}}{AUM_{i,t}} \quad (3)$$

where  $M$  denotes the total number of 401(k) retirement plans investing in fund  $i$  at time  $t$ , and  $AUM_{p,i,t}$  denotes the dollar amount invested by 401(k) plan  $p$  in fund  $i$  at the end of year  $t$ .  $IO_{i,t}^{401k}$  is hence a fund-specific, time-varying characteristic. Our first specification focuses on the demand function for the average stock  $n$ . Specifically, we estimate the AUM-weighted panel regression:

$$\log \left( \frac{w_{i,t}(n)}{w_{i,t}(0)} \right) = b_0 + \beta_{0,i} mb_t(n) + \beta_1' \mathbf{X}_t(n) + \beta_2 IO_{i,t}^{401k} + \alpha_t + \tilde{u}_{i,t}(n) \quad (4)$$

where each fund-stock holding observation in the panel is weighted by assets under management of fund  $i$ . The dependent variable represents the demand of stock  $n$  by fund  $i$  at time  $t$  with respect to the outside asset;  $mb_t(n)$  is the log market-to-book equity of firm  $n$  at time  $t$ ,  $\mathbf{x}_t(n)$  is the same vector of firm-specific characteristics specified in [Kojien and Yogo \(2019\)](#),  $IO_{i,t}^{401k}$  represents the fraction of fund  $i$ 's assets under management owned by 401(k) plans at time  $t$ , and  $\alpha_t$  are time (year) fixed effects. Note that  $IO_{i,t}^{401k}$  does not vary across stocks, but only across funds and over time. Similarly to [Kojien and Yogo \(2019\)](#), in this specification we pool all fund-stock holdings observations together, since very few funds in our sample holds more than 1,000 stocks at any given time.<sup>19</sup> Our coefficient of interest is  $\beta_2$ , representing the effect of 401(k) assets on fund  $i$  demand for the average stock  $n$ .

[Table 2](#) shows the results from the panel regression for the entire universe of funds (Panel A), mutual funds only (Panel B) and ETFs only (Panel C). Throughout, we use two-way (funds and time) clustered standard errors. Furthermore, in order to gauge the relative importance of the variables in the demand system, we standardize all variables to have unit standard deviation. Across specifications, the coefficient on fund-level 401(k) ownership,  $IO_{i,t}^{401k}$ , is positive and statistically significant for mutual funds (0.333,  $t$ -stat= 2.91) but small and insignificant for ETFs (0.011,  $t$ -stat= 0.11).<sup>20</sup> In terms of magnitude,

<sup>19</sup>We only have 268 fund-year observations, out of a total of 17,436, where the number of stocks is greater than 1,000.

<sup>20</sup>[Table A.1](#) reports the same results for AUM unweighted regressions. The coefficients for mutual funds

the coefficient of mutual funds on  $IO_{i,t}^{401k}$  ranks second in the set of characteristics after book equity.

The evidence that mutual funds strongly respond to fund-level 401(k) ownership while ETFs do not is interesting. It suggests that mutual funds exert more discretion in selecting their holdings based on fund-level ownership, relative to ETFs. This holds true despite the fact that mutual funds and ETFs display similar demand elasticity (approximately captured by the coefficient  $(1 - \beta_{0,i})$  on log market-to-book in Panel B and C of [Table 2](#)).<sup>21</sup> However, a word of caution is needed. An alternative interpretation could be inferred by [Figure 2](#): the fraction of aggregate ETF assets cumulatively owned by 401(k) is currently small, hovering around \$20-30bn, and, perhaps, not large enough to trigger a discernible demand shift. The last two rows of [Table 1](#) confirm this, with the average dollar amount across 401(k) plan assets invested in US equity mutual funds being around \$15mn, about four times the amount invested in ETFs. However, given the fast pace at which the investment of 401(k) plan in ETFs has been growing, results may differ in the future.

The results presented in [Table 2](#) mask some economically interesting trends. To this end, the left panel in [Figure 4](#) shows the evolution over time of the coefficient on fund-level 401(k) ownership for both mutual funds and ETFs.<sup>22</sup> The figure shows that the loading of mutual funds on fund-level 401(k) ownership is positive throughout the sample, and strongly increasing over time (with an average value of about 0.29, similar to that reported in Panel B of [Table 2](#)). In line with our previous discussion of a rapid growth of 401(k) allocation to ETFs, the effect of 401(k) ownership for ETF holdings becomes stronger over time, and marginally positive in the second part of our sample.

Panel A of [Table 3](#) reports GMM estimates of the non-linear version of equation (4). This allows us to take into account holdings of stocks that are in the fund investment universe, but not currently owned by the fund. To ease exposition, we display the results

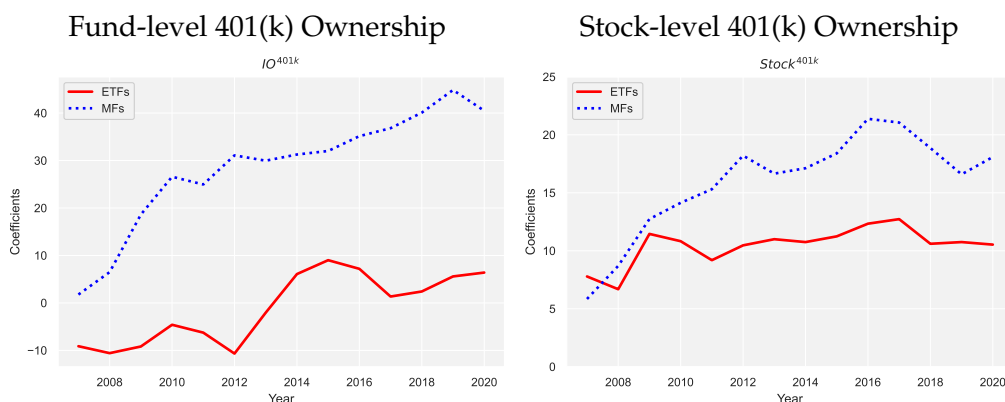
---

and ETFs are, respectively, 0.240 ( $t$ -stat= 4.29) and  $-0.060$  ( $t$ -stat=  $-1.10$ ), thus confirming our results.

<sup>21</sup>Recall that we consider only mutual funds and ETFs that are active by removing index mutual funds and ETFs.

<sup>22</sup>[Figure A.1](#) shows the coefficients on the other covariates. We find that the coefficients of ETFs and mutual funds on profitability, investment, and beta are similar. This is important since it highlights the economic significance of the observed difference in fund-level 401(k) ownership coefficients between mutual funds and ETFs.





**Figure 4: Coefficients on 401(k) ownership.** This figure shows the annual coefficient in equations (4) (fund level, left) and (6) (stock level, right) on 401(k) ownership, separately for mutual funds and ETFs, estimated by pooled OLS using assets under management as weights. The regression is estimated annually, and it includes fund-level fixed effect in the right panel. Variables are standardized (within each year) to make coefficients comparable. We multiply the coefficients on 401(k) ownership by 100, so that they can be interpreted as the percentage change in demand per one standard deviation change in the characteristic. The sample period is from 2007 through 2020.

only for mutual funds. The coefficient on fund-level 401(k) ownership is positive and statistically significant at the 10% level, even accounting for zero holdings.

## 4.2 Stock Level Channel

Next, we study the effect of 401(k) ownership at the individual stock level. This firm-specific characteristic, similar to, for example, book-to-market or beta, might be important in determining how much of a stock is demanded by funds through a feedback effect. In other words, fund managers may be more or less inclined to accumulate a position in a stock if they know it is largely owned by 401(k) plans. For example, if fund managers believe 401(k) allocations to funds to be stable,<sup>23</sup> and individual fund allocations do not drastically change over time,<sup>24</sup> then a larger stock ownership by 401(k) plans could signal potential stability in the stock investors' base as well. To this end, we calculate the fraction

<sup>23</sup>401(k) plans have persistent fund allocation in terms of proportions of assets under management, see [Section 2](#).

<sup>24</sup>This is consistent with the presence of investment mandates. For example, Table 1 in [Koijen and Yogo \(2019\)](#) reports that, across institutions, more than 82 percent of stocks currently held by an institution were also held in the previous quarter.

of stock  $n$  equity owned by 401(k) plans:

$$IO_t^{401k}(n) = \frac{\sum_{j=1}^I IO_{j,t}^{401k} * w_{j,t}(n) * AUM_{j,t}}{ME_t(n)} \quad (5)$$

where  $IO_{j,t}^{401k}$  is the fraction of fund  $j$  owned by all 401(k) at the end of year  $t$  (c.f., equation (3)),  $w_{j,t}(n)$  denotes the portfolio weight of equity fund  $j$  on stock  $n$  at the end of year  $t$ ,  $AUM_{j,t}$  denotes the assets under management (size) of fund  $j$ , and  $ME_t(n)$  is the market value of stock  $n$ . In other words, this variable represents the total ownership of stock  $n$  by 401(k) plans through both mutual funds and ETFs.

We then estimate the following panel regression specification:

$$\log \left( \frac{w_{i,t}(n)}{w_{i,t}(0)} \right) = b_0 + \beta_{0,i} mb_t(n) + \beta_1' \mathbf{X}_t(n) + \beta_2 IO_{-i,t}^{401k}(n) + \alpha_{i,t} + \tilde{u}_{i,t}(n) \quad (6)$$

where  $i$  represents a fund held in a 401(k) plan,  $IO_{-i,t}^{401k}(n)$  is the fraction of stock  $n$  cumulatively owned by 401(k) plans through funds, excluding that owned by fund  $i$ , and the other variables are defined as in Section 4.1. Note that by excluding investor  $i$  from the  $IO_t^{401k}(n)$  regressor in (5), we are studying how the portfolio choice of investor  $i$  is influenced by the stock-level 401(k) ownership derived from all *other* investors, thus reducing possible endogeneity concerns.<sup>25</sup>

Table 4 shows the results from the panel regression (6) for the entire universe of funds (Panel A), mutual funds (Panel B), and ETFs (Panel C). As in Section 4.1, we report two-way (funds and time) clustered standard errors, and fund-stock observations are AUM-weighted. Furthermore, to properly compare regression coefficients, we standardize all variables. Across specifications, the coefficient on stock-level 401(k) ownership,  $IO_t^{401k}(n)$ , is positive, it ranks second in terms of magnitude (among the characteristics included in  $\mathbf{X}_t(n)$ ), and it is statistically significant even after controlling for well known drivers of expected returns such as market beta, book-to-market, and profitability. This result highlights the relevance of stock-level 401(k) ownership as an important characteristic for fund allocation decisions. The coefficient for the universe of funds (0.149,  $t$ -

<sup>25</sup>Excluding investor  $i$  from the summation also addresses the concern that a stock owned only by one fund (a quite unlikely case) drives the results.

stat= 18.95, see Panel A) is mostly determined by mutual funds. Indeed, mutual funds display a loading on stock-level 401(k) ownership of 0.186, which is sixty percent greater than that of ETFs, equal to 0.115.<sup>26</sup>

The right panel in [Figure 4](#) shows the evolution of the coefficient on stock-level 401(k) ownership,  $IO_t^{401k}(n)$ , over time.<sup>27</sup> The coefficient is always larger and more volatile for mutual funds than for ETFs; in general, the magnitude of the coefficients are in line with the values reported in [Table 4](#). Panel B of [Table 3](#) reports estimates from the GMM estimation of the non-linear version of equation (6). The result shows a positive on stock-level 401(k) ownership of 0.25 – the second largest within the set of characteristics  $X_t(n)$  – statistically significant at the 1% level ( $t$ -stat: 25.49).

Before concluding, it is interesting to compare the results for ETFs in [Table 2](#) to those in [Table 4](#). Whereas a larger fund-level 401(k) ownership does not affect the demand for stocks – consistent with the idea that, for example, a growth ETF cannot take riskier bets and become a value ETFs – the stock-level 401(k)-ownership does, e.g., a growth ETF may pick stock A over stock B, despite having similar growth prospects and risk, simply because the fraction of stock A owned by 401(k) plans is larger. There are several reasons why ETFs could prefer stocks with large 401(k) ownership. On the one hand, it could be that the stock-level 401(k) ownership characteristic is correlated with the probability of the stock being included in the benchmarks tracked by ETFs (e.g., a growth index), since stocks with large institutional 401(k) ownership have already been screened and vetted by investors that bought them, and hence are included in benchmarks. On the other hand, it might be that some ETFs are not perfectly tracking their respective benchmarks, and hence have some leeway in choosing their stock holdings. Indeed, we find this to be the case. The top 30% of the ETFs in our sample display a large tracking error of more than 2.18% per year with respect to their benchmarks, suggesting that many “active” ETF

---

<sup>26</sup>[Table A.2](#) reports the same results without weighting the observations by the fund AUM. The coefficients for mutual funds and ETFs are 0.122 ( $t$ -stat= 11.07) and 0.082 ( $t$ -stat= 7.80), respectively, thus confirming a stronger effect for the former.

<sup>27</sup>[Figure A.2](#) shows the coefficients on the other covariates. Importantly, these coefficients are almost identical across a demand system specification that includes fund-level 401(k) ownership and the one that, instead, includes stock-level ownership. This suggests that the different behavior of ETFs toward fund- and stock-level ownership cannot be attributed to different specifications of the demand system.

managers have flexibility in selecting their stock holdings.

Next, we study the role of fund-level 401(k) ownership in determining the heterogeneous demand for stocks by fund managers in [Section 4.3](#) and the equilibrium price impact of stock-level 401(k) ownership in [Section 4.4](#).

### 4.3 Heterogeneous Demand for Stocks

The analysis so far shows that the amount of 401(k) assets managed by funds influence their average demand for stocks. In particular, the larger the 401(k) fund-level ownership, the stronger the demand for stocks, controlling for other prominent characteristics. However, the amount of 401(k) assets managed might not lift demand uniformly across stocks; rather, it may push fund managers toward certain types of stocks (e.g., winners) more than others (e.g., losers). In this section, we try to shed some light on the 401(k) asset-induced demand for specific stock characteristics.

We conjecture that the investment decisions of fund managers is related to the fraction of 401(k) assets they manage. Our first hypothesis is that funds controlling a larger fraction of 401(k) assets have preference for riskier assets, such as high-beta and momentum stocks.

**Hypothesis 1.** (*Relationship between 401(k) asset base and fund investments.*) *Funds managing more (sticky) 401(k) assets tend to invest in riskier assets (e.g., high-beta, momentum, and smaller stocks) given the risk of outflows from 401(k) plans is limited:*

$$H_1 : \beta_6 > 0, \text{ if } q = \text{mom, high-beta, small.}$$

Our second hypothesis is that the stability of the investor base allows funds to invest in stocks with embedded real options and long term growth prospects. Formally, we test for this hypothesis by studying the preference of fund managers for assets with long-duration of cashflows, which the literature has found to be less risky than short duration stocks.

**Hypothesis 2.** (*Funds with more 401(k) assets prefer longer-duration assets.*) *Funds managing more (sticky) 401(k) assets tend to invest in assets with longer-duration cashflows:*

$$H_2 : \beta_6 > 0, \text{ if } q = \text{long-duration}$$

To test our hypotheses, we first define the universe of stocks as the union of the investment universes of all equity funds in our sample.<sup>28</sup> Next, we unconditionally sort these stocks into five portfolios based on a given stock characteristic  $X_1$  (e.g., momentum). We then compute how much each fund  $i$  invest at time  $t$ , as a percentage of its assets, into the stocks within each quintile. This ensures that the fraction invested in all quintiles sums up to unity. As an example, a fund following a momentum strategy might invest 60% of its assets in stocks belonging to the top momentum quintile (“winners”) and 10% in each of the other four quintiles.

Furthermore, for each quintile sorted on a given characteristic, we also calculate the value of *other* lagged characteristics  $X_2, X_3, X_4, \dots$ , for that quintile. As an example, say that stocks A and B are the only two stocks in the winner portfolio (top quintile) at time  $t$ . We then calculate the book-to-market of the winner portfolio at time  $t - 1$  by value-weighting the book-to-market characteristic of stocks A and B. We focus on the characteristics implied by the [Fama and French \(2015\)](#) five factor model, to which we add momentum. Hence, in our momentum example,  $X_2$  is size,  $X_3$  book-to-market,  $X_4$  profitability,  $X_5$  investment, and  $X_6$  is the CAPM-beta.

We then estimate, using the “winner” portfolio as our running example, the following panel regression:

$$\begin{aligned} \%Share_{i,q,t+1} = & \beta_1 \times \beta_{q,t}^{CAPM} + \beta_2 \times BM_{q,t} + \beta_3 \times Prof_{q,t} + \beta_4 \times Inv_{q,t} + \beta_5 \times Size_{q,t} + \\ & + \beta_6 \times IO_{i,t}^{401k} + \text{controls} + u_{i,t} \quad (7) \end{aligned}$$

where  $\%Share_{i,q,t+1}$  is the fraction invested by fund  $i$  in the quintile  $q$  at time  $t + 1$ , and the time  $t$  predictors are the characteristics from the FF5 model ( $BM_{q,t}$ ,  $Prof_{q,t}$  and  $Inv_{q,t}$  are the book-to-market value, profitability, and investment rate of the winner quintile, and  $Size_{q,t}$  is the market capitalization of the stocks included in the winner quintile), and

---

<sup>28</sup>We do not consider the universe of stocks from, for example, the CRSP dataset, since most funds will have zero holdings for many stocks within that set.

$IO_{i,t}^{401k}$  is our variable of interest representing the fraction of mutual fund  $i$  owned by 401(k) pension plans. We also control for fund characteristics, namely fund size and the lagged fraction invested by fund  $i$  in the top quintile, and for fund family fixed effects.

We estimate the predictive regression (7) for portfolios sorted on (i) market-beta, one of the main characteristics the literature found to be important in fund managers' choices (Christoffersen and Simutin (2017) and Han et al. (2022)), (ii) momentum, (iii) size, and (iv) duration (computed as in Gormsen and Lazarus (2022)). In each instance, we exclude the left hand-side characteristic from the right-hand side predictors, e.g., if the dependent top quintile is "low-beta stocks", we do not include the lagged beta of the portfolio as a predictor. Importantly, this regression specification has on the left hand side the fraction invested in the quintile (defined by a specific characteristic) and, thus, it allows to determine the *portfolio* demand rather than the average individual stock demand, as in a standard demand-based regression framework.

Table 5 reports the results for the top quintile portfolio (Column 1), bottom quintile portfolio (Column 2), and their difference (Column 3). The first row reports our coefficient of interest,  $IO_{i,t}^{401k}$ , while the second row shows the coefficient on the lagged value of the portfolio share (i.e., the autoregressive coefficient of the dependent variable). Standard errors are reported below the coefficient estimates.

If a characteristic predicts returns with a negative sign (like size), then the bottom quintile contains large value for the characteristic (large stocks); in this way, a tilt toward the top portfolio and away from bottom one always captures an expected positive alpha.

First, we confirm the results by Christoffersen and Simutin (2017) that fund managers with large 401(k) ownership tend to increase their exposure to high-beta stocks (Panel A). Interestingly, we also observe a large decrease in their exposure to low-beta stocks.<sup>29</sup> Panel B in Table 5 documents a tilt away from short-duration stocks and, to a lesser extent, toward long-duration stocks, while Panel C and D show that managers take more risk (alpha) by tilting toward winners and away from large stocks, respectively. The tilts toward the smallest stocks or away from losers are insignificant, however.

---

<sup>29</sup>This is not mechanical. The increase in portfolio weights for the bottom portfolio could have come from a reduction in the middle quintiles.

These facts are interesting for several reasons. First, the tilt of fund managers' toward high-beta and long-duration stocks, and away from low-beta and short-duration stocks, could sustain the well known betting-against-beta (Frazzini and Pedersen, 2014) and duration anomalies (Weber, 2018). Similarly, the behavior of funds managing a large fraction of 401(k) assets tilting away from large stocks is consistent with the observed diminishing size premium. Second, the evidence for size and momentum suggest that fund managers try to improve not only relative returns (by investing in stocks with higher market beta), but they also care about absolute returns and alphas by attempting to reap the unconditional premium associated with size and momentum.

A natural question is whether the portfolio tilt implemented by mutual funds with large 401(k) ownership results in performances that beat the benchmarks. To this end, we estimate the fund relative returns and CAPM alphas as a function of lagged 401(k) ownership and other lagged fund characteristics, and report the results in Table 6. Column (2) shows that higher 401(k) ownership forecasts better performance relative to a style benchmark: a one-standard-deviation increase in 401(k) ownership increases relative performance by 194bps per year. In contrast, column (4) shows that higher 401(k) ownership is not associated with larger future alphas. This can be due to two effects. On the one hand, it is possible that the tilt toward winners and away from large stocks (e.g., positive alphas) is countervailed by the tilt toward high-beta and long-duration (e.g., negative alphas). Alternatively, it is plausible that the size of the tilts is not large enough to generate significant changes in alpha. Overall, a higher 401(k) ownership forecasts improved relative returns without a significant change in alpha, a result that is new to the literature. Interestingly, if pension plans care about relative returns more than absolute ones, then the relative outperformance documented in Panel A of Table 6 should have a positive effect on pension flows, triggering a potential feedback reaction, whereby 401(k) plans continue to invest more in those funds that beat their benchmarks (i.e., with better relative returns), which in turn happen to be those managing larger pension assets.<sup>30</sup>

---

<sup>30</sup>Christoffersen and Simutin (2017) provide anecdotal evidence (from investment policy statements of DC plans) that a large majority of DC plans list relative returns as the main criterion for investment.

## 4.4 Price Impact of 401(k) Plans

In this section we quantify the equilibrium price impact of a change in 401(k) stock-level ownership for firm  $n$ , accounting for the trading of all investors. Specifically, we estimate

$$\frac{\partial p_t(n)}{\partial IO_t^{401k}(n)} \quad (8)$$

where  $p$  is the log price of stock  $n$ . Following [Kojien and Yogo \(2019\)](#) and [Noh and Oh \(2020\)](#), this derivative can be computed analytically, at any time  $t$ , as the diagonal elements of the matrix  $\mathbf{M}$ :<sup>31</sup>

$$\mathbf{M} = \left( \mathbf{I} - \sum_i \beta_{0,i} A_i \mathbf{H}^{-1} \mathbf{G}_i \right)^{-1} \left( \sum_i \beta_{2,i} A_i \mathbf{H}^{-1} \mathbf{G}_i \right), \quad (9)$$

where we recall that  $\beta_{0,i}$  is the loading of investor  $i$  on market-to-book, and  $\beta_{2,i}$  is the coefficient on 401(k) ownership (see equation (1)). The matrices  $\mathbf{H} = \sum_{i=1}^I A_i \text{diag}(\mathbf{w}_i)$  and  $\mathbf{G}_i = \text{diag}(\mathbf{w}_i) - \mathbf{w}_i \mathbf{w}_i'$  instead do not depend on estimated parameters, but only on investors' weights  $\mathbf{w}$ . Finally,  $A_i$  denotes the asset under management of investor  $i$ .

The  $n$ -th diagonal entry of  $\mathbf{M}$ ,  $M_{n,n}$ , captures two effects. First, the matrix inside the inverse in equation (9) is the aggregate demand elasticity ([Kojien and Yogo, 2019](#)), and its diagonal elements are strictly positive when  $\beta_{0,i} < 1$  for all investors. If a firm is held by less price elastic investors, then the firm price will react more due to institutional demand for the  $IO_t^{401k}(n)$  characteristic. Second, the  $n$ -th diagonal entry of the matrix outside the inverse can be written as  $\frac{\sum_i \beta_{2,i} A_i w_i(n) (1 - w_i(n))}{\sum_i A_i w_i(n)}$ , and represents an AUM weighted average of the coefficients on the 401(k) stock-level ownership (multiplied by  $1 - w_i(n)$ ). This implies that the price pressure is larger if a firm faces owners that are large and exhibit a high coefficient on the  $IO_t^{401k}(n)$ . In other words, the institutional price pressure that a given firm  $n$  receives due to a change in the level of 401(k) ownership is a weighted average of  $IO_t^{401k}(n)$  coefficients of its institutional owners, adjusted for their demand

<sup>31</sup>To compute this expression one has to exploit the identity  $\mathbf{p} = \log(\sum_i A_i \mathbf{w}_i) - \mathbf{s}$  (where  $\mathbf{s}$  denotes the vector of shares outstanding) which holds by market clearing. See Appendix A in [Noh and Oh \(2020\)](#) for additional details.



elasticities.

To compute the price impact  $M_{n,n}$  we need to consider the entire investor universe, i.e., not only mutual funds and ETFs. To this end, we use data on institutional common stock holdings from the Thomson Reuters Institutional Holdings Database (s34 file). We follow the [Kojien and Yogo \(2019\)](#) classification of institutions into six types (i.e.,  $i = 1, \dots, 6$ ): banks, insurance companies, investment advisors, mutual funds, pension funds, and other 13F institutions. We recall that the s34 file provides a different level of granularity relative to our analysis in [Section 4.1](#) and [Section 4.2](#), since it reports aggregate holdings at the investor level (e.g., for all funds managed by Fidelity). In Internet Appendix [B.1](#) we confirm that stock-level 401(k) ownership remains an important characteristic for fund allocation decisions when we use the s34 dataset instead of Morningstar.<sup>32</sup>

[Figure 5](#) displays the two key ingredients required to compute the price impact: the coefficient on book-to-market driving demand elasticities (Panel A) and the coefficient on 401(k) stock-level ownership (Panel B) for each of the six groups of investors. These coefficients are estimated year-by-year using GMM, accounting for zero holdings, under moment condition (2).<sup>33</sup> We confirm the results in [Kojien and Yogo \(2019\)](#) that mutual funds have less elastic demand than investment advisors for most of the sample period, and that insurance companies and pension funds have become less elastic over time. The coefficient in Panel B captures institutional demand for 401(k) stock-level ownership. When positive, it implies that investor  $i$  allocates at time  $t$  more weight to stocks with higher 401(k) ownership, controlling for other stock characteristics. We see that mutual funds, banks and insurance companies tilt their portfolio toward stocks with high-level of 401(k) ownership more than other types of institutions. In contrast, investment advisors do not manifest such a tilt. Interestingly, the tilt of pension funds toward stocks with high level of  $IO_t^{401k}(n)$  increases over our sample period suggesting an intricate relation between the sample of funds offered by 401(k) plans, their holdings, 401(k) plan investor preferences, and the type of individual stocks preferred by pension plans (e.g., green stocks).

---

<sup>32</sup>Mindful of potential gaps in coverage of institutional holdings, in Internet Appendix [B.2](#) we validate our price impact results by replacing s34 data with data on 13F filings from [Backus et al. \(2021\)](#).

<sup>33</sup>To obtain the price impact, we estimate the coefficients institution by institution (or by group if the holdings are small). Variables are standardized within each institution (or group) and for each year.

Finally, the evidence in Panel B for investors other than mutual funds emphasizes the relevance of stock-level 401(k) ownership as an important characteristic while further alleviating exogeneity concerns: we are using holdings of banks, insurance, etc. as left hand side variable, while only employing holdings of mutual funds in the construction of our stock-level 401k ownership (right hand side variable).

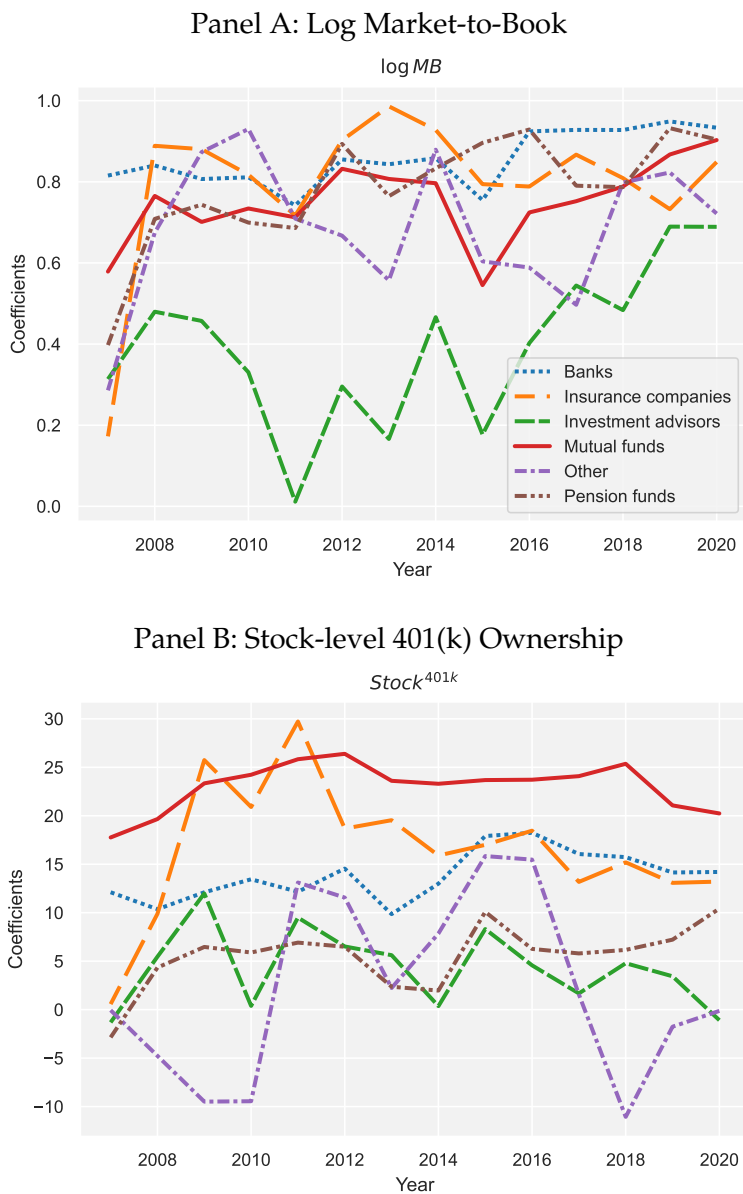
Given estimates of  $\beta_{0,i,t}$  and  $\beta_{2,i,t}$ , we can calculate, at every time period  $t$ , the firm-level institutional pressure with respect to 401(k) ownership. The top left panel in [Figure 6](#) shows the cross-sectional distribution of price impact across stocks. The aggregate price impact for the median stock (solid black line) has generally increased over time, and the cross-sectional spread has also significantly expanded over our period. The stronger effect over time can be related to the shift from active to passive investing of the last decade, since equation (9) highlights that the presence of more inelastic investors results in larger price pressure. A one standard deviation increase in 401(k) ownership, around 1% in 2007 and 2% in 2020, leads to a price impact (for the median stock) slightly less than 20 percent in 2007 and of about 60 percent in 2020. The remaining panels display the aggregate price impact for extreme quintile portfolios of stocks sorted on book-to-market (top right panel), market beta and size (bottom left and right panels, respectively). We observe that the average price impact has increased for large stocks with a sharp jump in 2015, while it has remained relatively stable for small stocks. This resonates well with [Haddad et al. \(2022\)](#) who find that investor elasticities are lower for larger stocks (i.e., investors are more reluctant to change their positions for large stocks than for small stocks), given tracking error concerns.<sup>34</sup>

We do not observe noticeable differences for stocks sorted on book-to-market or betas, which suggests that our 401(k) ownership provides additional information that is not subsumed by well-known stock characteristics. For book-to-market and betas-sorted portfolios, we again observe a positive low-frequency trend of price impact from 2008 to 2020. However, we also observe an interesting cyclical pattern around the low-frequency trend, particularly for value and high-beta stocks. This variation over time must be driven

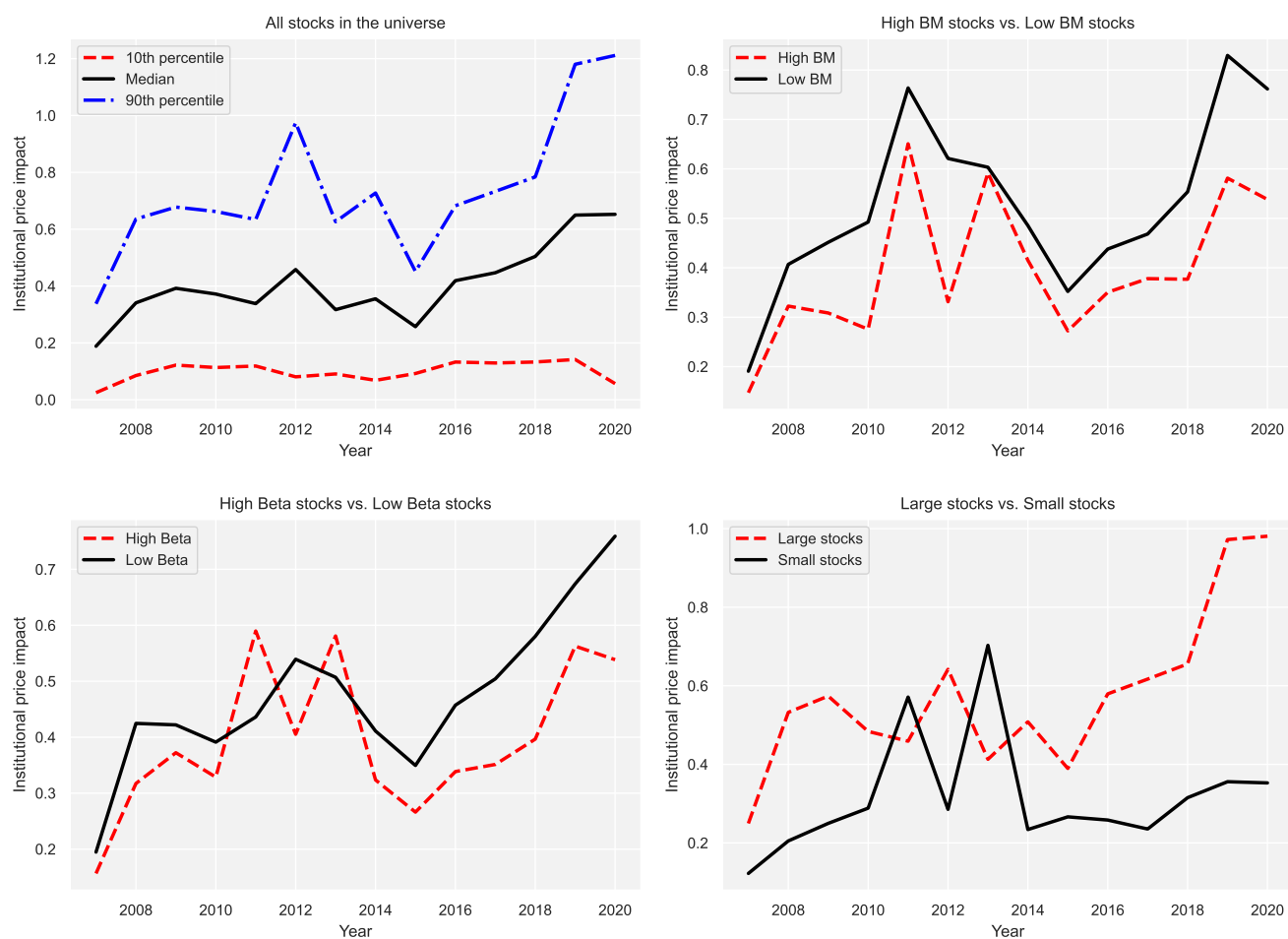
---

<sup>34</sup>In the U.S. stock market, large corporations like Apple make up a substantial fraction of total market capitalization and, as a consequence, a change in portfolio weight would have a large effect on an institution's portfolio return.

by shifts in investor's preferences (the coefficient  $\beta_{2,i,t}$ ) beyond those associated with recessionary and expansionary periods.



**Figure 5: Price impact: relevant coefficients.** This figure shows the annual coefficient on log market-to-book (top panel) and stock-level 401(k) ownership (bottom panel) for financial institutions in Thomson Reuters holding (s34) estimated annually by GMM with zero weights. Variables are standardized (within each year) to make coefficients comparable. We report the cross-sectional mean of the estimated coefficients by institution type, weighted by assets under management. The coefficient on 401(k) ownership is multiplied by 100. The sample period is from 2007 through 2020.



**Figure 6: Institutional price impact.** This figure shows the price impact of a change in the stock-level 401(k) ownership estimated through the diagonal elements of the matrix  $\mathbf{M}$  defined in (9). The top left panel shows the 10th percentile, median and 90th percentile of price impact across all stocks. The remaining panels plot the average price impact across the stocks in the top quintile and bottom quintile of portfolios sorted on (top right), beta (bottom left), and size (bottom right), using NYSE break points as cutoffs.

## 5 Conclusion

In this paper, we study the impact of 401(k) pension plans ownership on fund demand for individual stocks. More precisely, we estimate a demand system linking 401(k) plans ownership of both funds and individual stocks to the quantity and type of stocks demanded by funds.

We hypothesize that 401(k) allocations can affect stock demand through two channels. The first channel through which 401(k) allocations influence stock demand is by direct flows to mutual funds and ETFs which, in turn, use that additional liquidity to increment their equity exposure. We label this the *fund level* channel. The second channel through which 401(k) plans affect the demand of individual stocks is by ownership of the stock itself. The fraction of an individual stock owned by 401(k) plans is an additional stock characteristic, similarly to book-to-market or momentum, that fund managers may take into account when evaluating how many shares of a specific stock to purchase. We label this the *stock level* channel.

We find strong support for the first channel only for mutual funds, and for the second channel for both mutual funds and ETFs.

First, with regard to the fund-level channel, we find that mutual fund managers take into account the amount of 401(k) assets they manage when deciding their portfolio allocation. Specifically, mutual funds managing a higher fraction of 401(k) assets display larger demand for stocks. The insignificant result for ETFs suggest that mutual fund managers have more discretion than ETFs ones.

Second, with respect to the stock-level channel, we find that the amount of company shares owned by 401(k) plans is an important characteristic in explaining the decision of mutual funds and ETFs to invest in a stock. For a one standard deviation increase in 401(k) stock ownership, the average mutual fund demands approximately 43.3% ( $t$ -stat: 14.57) more of the stock. Differently from the fund-level case, the average ETF also increases exposure to the stock by 15.5% for each standard deviation increase in 401(k) stock ownership. The result for ETFs might appear surprising considering that they are often believed to be passive funds. However, we find that 30% of the ETFs in our sample

display a tracking error of more than 2.18% per year with respect to their benchmarks, suggesting that many ETF managers still have flexibility in selecting their stock holdings.

Third, we shed light on the 401(k) ownership-induced fund demand for specific stock characteristics and fund performances. We find that fund's demand for stocks is heterogeneous as a function of 401(k) fund-level ownership. Funds managers with a larger fraction of 401(k) assets prefer stocks that are winners, and have high-beta and long-duration. The fact that managers tilt their portfolio toward long-duration and high-beta stocks, and away from stocks with short-duration and low-beta, can reinforce observed pricing anomalies. We also document that funds with higher 401(k) ownership achieve higher returns relative to their benchmarks, without influencing fund alphas, consistent with the theoretical literature on benchmarking and fund managers incentives.

Lastly, we estimate the equilibrium price impact of a change in the stock-level 401(k) ownership for the cross-section of stocks, every year, and find the institutional price pressure to be positive and increasing over time. For the median stock, it increases from around 0.2 in 2007 to 0.6 in 2020, and hovers around 0.8 over our sample for stocks lying on the 90th percentile of the price impact distribution. The positive trend of price impact, which increases almost monotonically between 2008 and 2020 is consistent with the shift from active to passive investing of the last decade documented in the literature ([Kojien et al., 2022](#)).

Overall, our results suggest that pension assets are a key determinant of asset allocation decisions and stock demand of both mutual fund and ETF managers. The key contribution of this paper is to quantify such effects.

## References

- Backus, Matthew, Christopher Conlon, and Michael Sinkinson (2021) "Common Ownership in America: 1980–2017," *American Economic Journal: Microeconomics*, Vol. 13, No. 3, pp. 273–308, August.
- Baker, Malcolm, Brendan Bradley, and Jeffrey Wurgler (2011) "Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly," *Financial Analysts Journal*, Vol. 67, No. 1, pp. 1397–1422, August.
- Bretscher, Lorenzo, Lukas Schmid, Ishita Sen, and Varun Sharma (2021) "Institutional Corporate Bond Pricing," *Working Paper*.
- Buffa, Andrea M., Dimitri Vayanos, and Paul Woolley (2022) "Asset Management Contracts and Equilibrium Prices," *Journal of Political Economy*, Vol. 130, No. 12, pp. 000–000.
- Christoffersen, Susan E. K. and Mikhail Simutin (2017) "On the Demand for High-Beta Stocks: Evidence from Mutual Funds," *The Review of Financial Studies*, Vol. 30, No. 8, pp. 2596–2620, 04.
- Dou, Winston Wei, Leonid Kogan, and Wei Wu (2022) "Common Fund Flows: Flow Hedging and Factor Pricing," *Working Paper*.
- Egan, Mark, Alexander MacKay, and Hanbin Yang (2021) "What Drives Variation in Investor Portfolios? Evidence from Retirement Plans," *Working Paper*.
- Fama, Eugene F and Kenneth R French (2015) "A five-factor asset pricing model," *Journal of Financial Economics*, Vol. 116, No. 1, pp. 1–22.
- Frazzini, Andrea and Lasse Pedersen (2014) "Betting against beta," *Journal of Financial Economics*, Vol. 111, No. 1, pp. 1–25.
- Gabaix, Xavier, Ralph S. J. Koijen, Federico Mainardi, Sangmin Oh, and Motohiro Yogo (2022) "Asset Demand of U.S. Households," *Working Paper*.

- Gormsen, Niels and Eben Lazarus (2022) "Duration Driven Returns," *Journal of Finance, Forthcoming*, 6.
- Haddad, Valentin, Paul Huebner, and Erik Loualiche (2022) "How Competitive is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing," *Working Paper*.
- Han, Xiao, Nikolai Roussanov, and Hongxun Ruan (2022) "Mutual Fund Risk Shifting and Risk Anomalies," *Working Paper*.
- Koijen, Ralph S. J. and Motohiro Yogo (2019) "A Demand System Approach to Asset Pricing," *Journal of Political Economy*, Vol. 127, No. 4, pp. 1475 – 1515.
- (2020) "Exchange Rates and Asset Prices in a Global Demand System," NBER Working Papers 27342, National Bureau of Economic Research, Inc.
- Koijen, Ralph S.J., Robert J. Richmond, and Motohiro Yogo (2022) "Which Investors Matter for Equity Valuations and Expected Returns?" *Working Paper*.
- Noh, Don and Sangmin Oh (2020) "Measuring institutional pressure for greenness: A demand system approach," NBER Working Papers 1291, National Bureau of Economic Research, Inc.
- Pástor, L'uboš, Robert F. Stambaugh, and Lucian A. Taylor (2015) "Scale and skill in active management," *Journal of Financial Economics*, Vol. 116, No. 1, pp. 23–45, apr.
- Schwarz, Christopher G. and Mark E. Potter (2016) "Revisiting Mutual Fund Portfolio Disclosure," *Review of Financial Studies*, Vol. 29, No. 12, pp. 3519–3544, sep.
- Sialm, Clemens and Laura Starks (2012) "Mutual Fund Tax Clienteles," *Journal of Finance*, Vol. 67, No. 4, pp. 1397–1422, August.
- Sialm, Clemens, Laura T. Starks, and Hanjiang Zhang (2015) "Defined Contribution Pension Plans: Sticky or Discerning Money?" *The Journal of Finance*, Vol. 70, No. 2, pp. 805–838.



Weber, Michael (2018) "Cash flow duration and the term structure of equity returns,"  
*Journal of Financial Economics*, Vol. 128, No. 3, pp. 486–503.

	25th	Median	Mean	75th
$IO^{401k}$ (all funds)	0.31%	2.72%	8.05%	10.46%
$IO^{401k}$ (MFs only)	0.86%	3.87%	9.25%	12.59%
$IO^{401k}$ (ETFs only)	0.02%	0.05%	0.89%	0.21%
Persistence on fund allocation	0.27	0.59	0.54	0.82
Total assets of 401(k) plans (in \$ millions)	4.34	10.73	91.57	30.13
Allocation in US equity index funds (in \$ millions)	0.36	1.28	9.76	4.39
Allocation in US equity mutual funds (in \$ millions)	0.83	2.78	14.98	8.44
Allocation in US ETFs (in \$ millions)	0.11	0.45	3.84	1.64

**Table 1: Summary statistics.** This table reports summary statistics of the cross-sectional distribution of 401(k) plans characteristics.  $IO^{401k}$  indicates the fraction of a fund assets owned collectively by 401(k) plans. Persistence on fund allocation is the AR(1) coefficient on the fraction of 401(k) plan assets invested in a specific fund. Total assets are the total net assets of 401(k) plans. Index funds include mutual funds classified according to the Morningstar variables "index funds" and "enhanced index", and ETFs with the S&P500 index as benchmark. Annual data from 2007 to 2020.

<b>Panel A: All Funds</b>			
	Coefficient	s.e.	<i>t</i> -stat
$IO_{i,t}^{401k}$	0.169**	0.071	2.390
Log market-to-book	0.766***	0.104	7.340
Log book equity	1.289***	0.091	14.170
Operating profitability	0.036**	0.013	2.800
Beta	-0.033**	0.012	-2.660
Investment	-0.013	0.014	-0.940
Dividend-to-book	-0.047***	0.013	-3.600
<b>Panel B: Mutual Funds</b>			
	Coefficient	s.e.	<i>t</i> -stat
$IO_{i,t}^{401k}$	0.333***	0.114	2.910
Log market-to-book	0.637***	0.067	9.510
Log book equity	1.032***	0.093	11.050
Operating profitability	0.031*	0.017	1.810
Beta	-0.042***	0.014	-2.910
Investment	-0.021	0.016	-1.280
Dividend-to-book	-0.090***	0.015	-5.970
<b>Panel C: ETFs</b>			
	Coefficient	s.e.	<i>t</i> -stat
$IO_{i,t}^{401k}$	0.011	0.098	0.110
Log market-to-book	0.560***	0.123	4.540
Log book equity	1.105***	0.111	9.930
Operating profitability	0.042**	0.016	2.660
Beta	-0.064***	0.018	-3.520
Investment	-0.031*	0.015	-2.030
Dividend-to-book	0.035	0.023	1.530

**Table 2: Demand system estimation - Fund level  $IO_{i,t}^{401k}$ .** This table reports estimates of the panel regression

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = b_0 + \beta_{0,i}\widehat{mb}_t(n) + \beta_1'X_t(n) + \beta_2 IO_{i,t}^{401k} + \alpha_t + \bar{u}_{i,t}(n)$$

where  $\widehat{mb}_t(n)$  is the instrumented log market equity-to-book equity, and  $X_t(n)$  includes the same variables as in [Kojen and Yogo \(2019\)](#), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio.  $IO_{i,t}^{401k}$  is the 401K plans ownership of fund  $i$ , and  $\alpha_t$  are time fixed effects. The funds in the regressions are AUM-weighted. Panel A reports results using all funds (including index mutual funds and ETFs), Panel B only mutual funds, and Panel C only ETFs. In Panels B and C we exclude index funds. Variables are standardized. Standard errors are double clustered by fund and time. The sample period is from 2007 to 2020. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

<b>Panel A: Fund Level 401(k) Ownership (mutual funds)</b>			
	Coefficient	s.e.	t-stat
$IO_{i,t}^{401k}$	0.048*	0.027	1.770
Log market-to-book	0.199***	0.036	5.570
Log book equity	0.530***	0.027	19.780
Operating profitability	0.051***	0.007	6.980
Beta	-0.131***	0.009	-15.020
Investment	-0.094***	0.007	-13.340
Dividend-to-book	-0.037***	0.008	-4.730

<b>Panel B: Stock Level 401(k) Ownership (mutual funds)</b>			
	Coefficient	s.e.	t-stat
$IO_{stock,t}^{401k}$	0.250***	0.010	25.490
Log market-to-book	0.132***	0.034	3.890
Log book equity	0.590***	0.025	23.750
Operating profitability	0.067***	0.007	9.280
Beta	-0.124***	0.008	-14.720
Investment	-0.097***	0.007	-13.800
Dividend-to-book	0.014*	0.008	1.800

**Table 3: Demand system estimation - GMM with fund- and stock-level 401(k) ownership.** This table reports GMM estimates of the regression

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = \exp \left\{ b_0 + \beta_{0,i} \widehat{mb}_t(n) + \beta_1' X_t(n) + \beta_2 IO_t^{401k} + \alpha_t \right\} \tilde{u}_{i,t}(n)$$

where  $\widehat{mb}_t(n)$  is the instrumented log market equity-to-book equity, and  $X_t(n)$  includes the same variables as in [Kojien and Yogo \(2019\)](#), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio.  $IO_t^{401k}$  indicates the 401(k) plans ownership of fund  $i$  (Panel A) or of the individual stock  $n$  excluding the effect through fund  $i$  (Panel B). Panel A (Panel B) reports fund-level (stock-level) results for mutual funds. The estimation includes observations of mutual funds with zero stock-holdings but still in the investment universe, and observations are AUM-weighted. Variables are standardized. Standard errors are double clustered by fund and time. The sample period is from 2007 to 2020. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

<b>Panel A: All Funds</b>			
	Coefficient	s.e.	t-stat
$IO_t^{401k}(n)$	0.149***	0.008	18.950
Log market-to-book	0.812***	0.059	13.770
Log book equity	1.411***	0.054	26.140
Operating profitability	0.030***	0.009	3.380
Beta	-0.004	0.005	-0.930
Investment	0.011**	0.005	2.170
Dividend-to-book	0.000	0.006	0.070
<b>Panel B: Mutual Funds</b>			
	Coefficient	s.e.	t-stat
$IO_t^{401k}(n)$	0.186***	0.017	11.160
Log market-to-book	0.422***	0.085	5.000
Log book equity	0.991***	0.085	11.670
Operating profitability	0.072***	0.017	4.200
Beta	-0.020**	0.008	-2.670
Investment	0.009	0.011	0.760
Dividend-to-book	-0.014	0.010	-1.390
<b>Panel C: ETFs</b>			
	Coefficient	s.e.	t-stat
$IO_t^{401k}(n)$	0.115***	0.011	10.210
Log market-to-book	0.964***	0.044	22.090
Log book equity	1.515***	0.025	59.930
Operating profitability	0.017**	0.007	2.500
Beta	0.004	0.006	0.700
Investment	0.013**	0.006	2.200
Dividend-to-book	0.003	0.009	0.390

**Table 4: Demand system estimation - Stock level  $IO_t^{401k}(n)$ .** This table reports estimates of the panel regression

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = b_0 + \beta_{0,i}\widehat{mb}_t(n) + \beta_1'X_t(n) + \beta_2 IO_{-i,t}^{401k}(n) + \alpha_{i,t} + \bar{u}_{i,t}(n)$$

where  $\widehat{mb}_t(n)$  is the instrumented market-to-book equity, and  $X_t(n)$  includes the same variables as in [Kojien and Yogo \(2019\)](#), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio.  $IO_{-i,t}^{401k}(n)$  is the 401K plans ownership of stock  $n$  (excluding the effect through investor  $i$ ), and  $\alpha_{i,t}$  are fund-by-time fixed effects. The funds in the regressions are AUM-weighted. Panel A reports results using all funds (including index mutual funds and ETFs), Panel B only mutual funds, and Panel C only ETFs. In Panels B and C we exclude index funds. Variables are standardized. Standard errors are double clustered by fund and time. The sample period is from 2007 to 2020. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

<b>Panel A: Beta</b>			
	High (Low Beta)	Low (High Beta)	High-Low
	(1)	(2)	(3)
$IO_{i,t}^{401k}$	-0.027*** (0.006)	0.015*** (0.004)	-0.031*** (0.009)
$weight_t$	0.883*** (0.020)	0.872*** (0.016)	0.897*** (0.015)
<b>Panel B: Duration</b>			
	High (Short Duration)	Low (Long Duration)	High-Low
	(1)	(2)	(3)
$IO_{i,t}^{401k}$	-0.028*** (0.007)	0.008** (0.003)	-0.023*** (0.006)
$weight_t$	0.890*** (0.009)	0.863*** (0.011)	0.900*** (0.009)
<b>Panel C: Momentum</b>			
	High (Winner)	Low (Losers)	High-Low
	(1)	(2)	(3)
$IO_{i,t}^{401k}$	0.021** (0.010)	-0.006 (0.006)	0.028** (0.014)
$weight_t$	0.869*** (0.039)	0.808*** (0.019)	0.859*** (0.034)
<b>Panel D: Size</b>			
	High (Small)	Low (Large)	High-Low
	(1)	(2)	(3)
$IO_{i,t}^{401k}$	0.002 (0.005)	-0.015** (0.007)	0.014 (0.008)
$weight_t$	0.930*** (0.023)	0.996*** (0.003)	1.000*** (0.005)

**Table 5: Effect of 401(k) ownership on the types of stocks preferred by funds.** This table reports results from regressions of the fraction of assets invested by mutual funds in a given portfolio in year  $t + 1$ ,  $w_{i,t+1}^P$  for  $P = High, Low, High - Low$ , on the 401(k) plans ownership of fund  $i$ ,  $IO_{i,t}^{401k}$ , controlling for lagged portfolio weights,  $w_{i,t}^P$ , as well as for the value-weighted characteristics of the portfolio, fund size at the end of year  $t$ , and fund family fixed effects. The portfolio characteristics are log market equity, log market-to-book, operating profitability, stock market beta, asset growth, and past 12-month returns, where we exclude the variable from the regressors when it is used as dependent variable. From left to right, the columns report the top and the bottom quintiles, and their differences. Standard errors are reported in brackets below the coefficients. Annual data from 2007 through 2020. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

	Relative return $_{t+1}$		Market $\alpha_{t+1}$	
	(1)	(2)	(3)	(4)
Fund $\beta_t$	1.918*** (2.755)	1.851** (2.497)	-0.458*** (-6.472)	-0.465*** (-5.976)
Log fund size $_t$	0.065 (0.995)	0.040 (0.617)	0.029*** (4.089)	0.030*** (3.623)
Expenses $_t$	0.742** (2.196)	0.798** (2.350)	0.072** (2.171)	0.074* (1.681)
Relative return $_t$	-0.014 (-0.621)	-0.015 (-0.657)	0.000 (0.263)	0.000 (0.238)
Turnover $_t$	-0.223 (-1.595)	-0.223 (-1.594)	0.003 (0.110)	0.003 (0.109)
Amihud illiquidity $_t$	0.052 (1.156)	0.057 (1.218)	0.001 (0.231)	0.001 (0.283)
$IO_{i,t}^{401k}$		1.938*** (3.000)		0.095 (0.984)

**Table 6: Fund performance and fund-level 401(k) ownership.** This table reports estimates of yearly panel regressions of measures of mutual fund performance on various lagged fund characteristics. Columns (1)-(2) report results using the fund relative return (e.g., the difference between the annual fund return and Morningstar category benchmark) as dependent variable, while columns (3)-(4) use the fund CAPM-alpha. Fund  $\beta$  is estimated from a monthly CAPM regression each year. Log fund size is the logarithm of the fund AUM. Expenses is net expense ratio, which is the total net expenses divided by the fund's average net assets. Turnover is a measure of the fund's trading activity, which is computed by taking the lesser of purchases or sales (excluding all securities with maturities of less than one year) and dividing by average monthly net assets. Amihud illiquidity is the value-weighted average of individual stock illiquidity based on the market value of stocks in the fund. Individual stock illiquidity is defined as the past 12-month average of its daily absolute return scaled by dollar volume.  $IO_{i,t}^{401k}$  is the fraction of fund assets owned by 401(k) plans. Regressions include year and fund family fixed effect.  $t$ -statistics (standard errors clustered by fund) are presented in parentheses.

## Internet Appendix A

### A.1 Coefficients on Other Characteristics

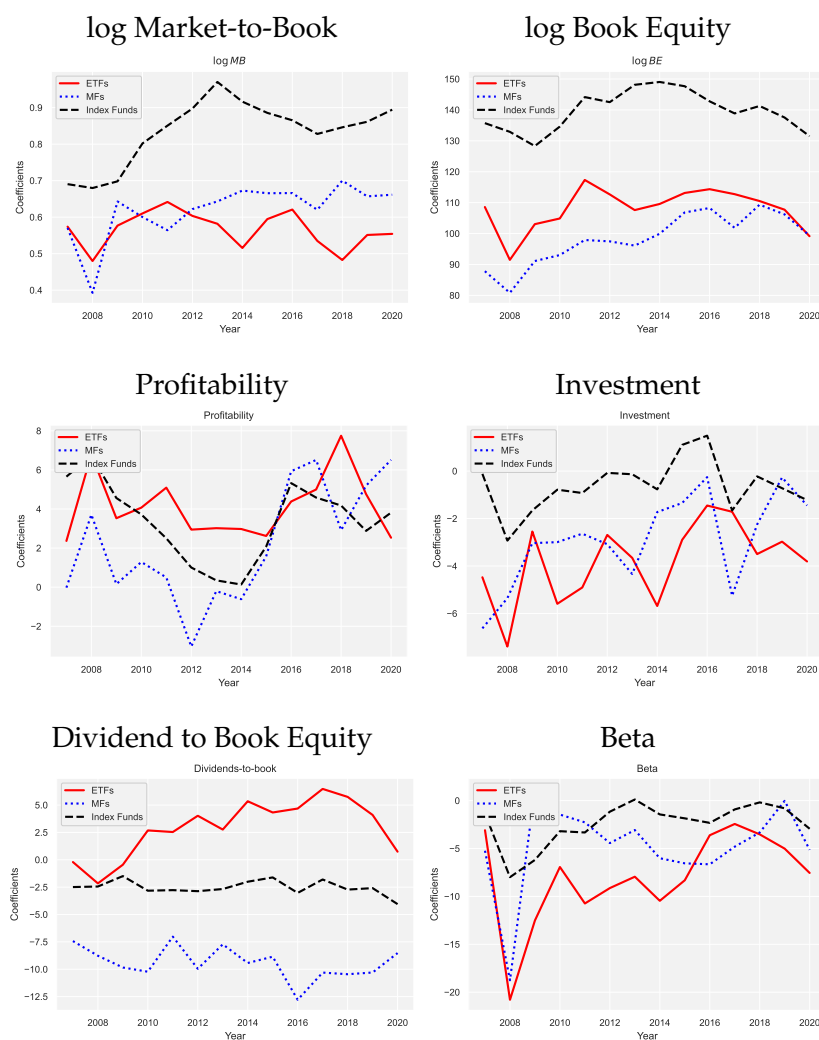
Figure A.1 shows annual estimates of the coefficients on market-to-book and characteristics for mutual funds (blue dotted line) and ETFs (red solid line) for the demand system that includes fund-level 401(k) ownership (see equation (4)). The coefficient on fund-level 401(k) ownership is displayed in the left panel of Figure 4.

To validate our estimation, we also report the coefficient estimates for index (mutual and ETF) funds (c.f., Section 2). If the estimation of our characteristics-based demand system is valid, one should recover a unit coefficient on log market equity, and zero on the other characteristics for an hypothetical index fund. Albeit the coefficient on market equity (which can be obtained from the coefficient on log market-to-book equity and log book equity) is not exactly one, we still notice that index funds are inelastic, and substantially more so than active mutual funds and ETFs. Furthermore, the coefficient of index funds on other characteristics is close to zero, the sole exception being the dividend-to-book equity. Thus we confirm the validity of our characteristics-based demand estimation and of our criteria to categorize index funds.

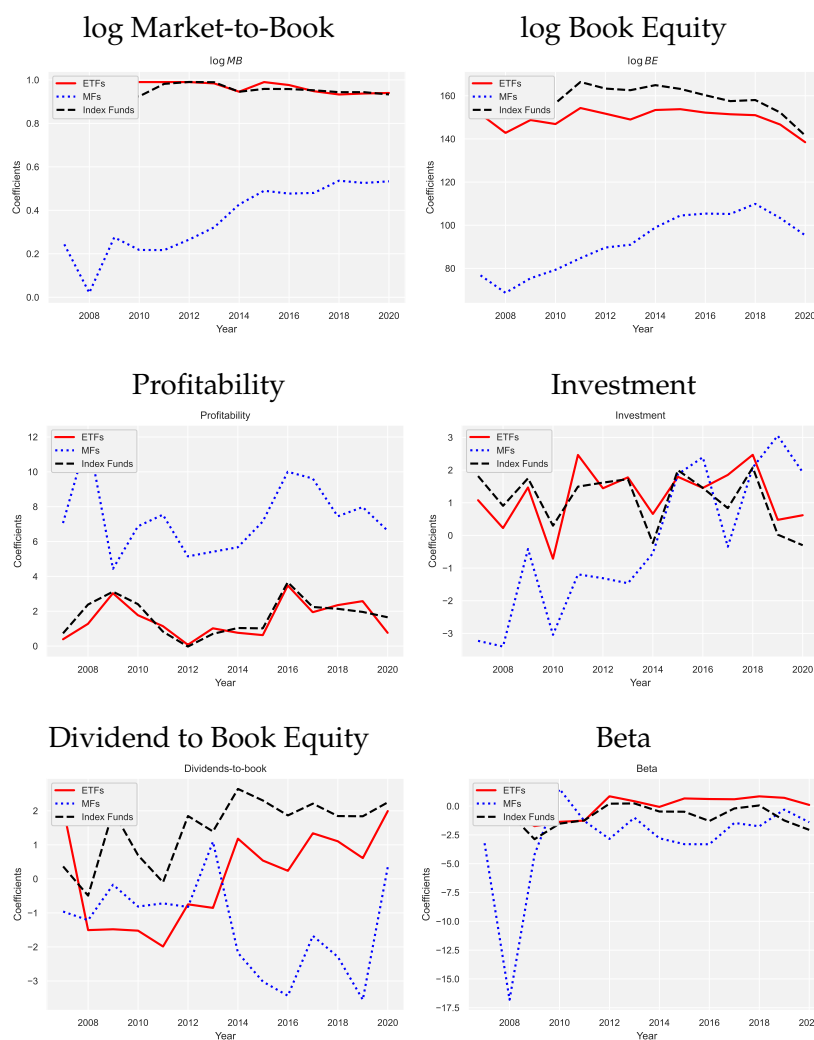
We also observe that, with the sole exception of dividend-to-book, ETFs and mutual funds display very similar coefficients on other prominent characteristics like betas, profitability and investment. This makes even more striking the large difference on 401(k) ownership loadings between mutual funds and ETFs documented in Figure 4.

Figure A.2 shows annual estimates of the coefficients on market-to-book and the other characteristics for the demand system that includes stock-level 401(k) ownership (see equation (6)). Comparing Figure A.1 to Figure A.2, we see that the coefficients are almost identical across the two specifications. In particular, demand elasticity is almost unaffected in terms of magnitude and time variation by the inclusion of stock-level and exclusion of fund-level ownership in the demand system. This is comforting because it suggests that the different behavior of ETFs toward fund- and stock-level ownership cannot be attributed to changes originating from different demand system specifications (namely equations (4) and (6)).





**Figure A.1: Coefficients on the other characteristics - Fund level.** This figure shows the annual coefficients in (4), separately for mutual funds, ETFs, and index funds, estimated by pooled OLS using assets under management as weights. The regression is estimated year by year. Except for log market-to-book equity, we standardize characteristics (within each year) and multiply the coefficients by 100, so that they can be interpreted as the percentage change in demand per one standard deviation change in the characteristic. The sample period is from 2007 through 2020.



**Figure A.2: Coefficients on the other characteristics - Stock level.** This figure shows the annual coefficients in (6), separately for mutual funds, ETFs, and index funds, estimated by pooled OLS using assets under management as weights. The regression is estimated year by year. Except for log market-to-book equity, we standardize characteristics (within each year) and multiply the coefficients by 100, so that they can be interpreted as the percentage change in demand per one standard deviation change in the characteristic. The sample period is from 2007 through 2020.

## A.2 Exogeneity of fund-level $IO_{i,t}^{401k}$

In this section, we conduct robustness tests to diffuse endogeneity concerns related to the fund-level 401(k) ownership variable  $IO_{i,t}^{401k}$ .

Equation (1) is estimated at the *stock* level, i.e., the individual stock demand by funds as a function of stock characteristics. Hence, perhaps, our  $IO_{i,t}^{401k}$  variable could be endogenous if other *stock-specific characteristics* affecting the behavior of mutual fund managers, other than those explicitly used as regressors in our model, are correlated with our variable. We believe that the fraction of a fund owned by 401(k) plans (e.g., our fund-level  $IO_{i,t}^{401k}$  variable) is likely orthogonal to any other stock-specific characteristic not explicitly controlled for in the model (e.g., the regressors used in the [Kojien and Yogo \(2019\)](#) framework). As an example, the number of employees of Exxon Mobil is arguably orthogonal to the amount of money CalPers decides to invest in any BlackRock mutual fund.

However, there could be *fund-specific* characteristics correlated with the 401(k) ownership of a fund. For example, 401(k) plans might prefer to invest in larger mutual funds or ETFs. Among many potential fund-specific characteristics, 401(k) pension plans are likely selecting funds based on the following fund characteristics ([Christoffersen and Simutin \(2017\)](#)): the fund strategy or style (e.g., “growth”), the investment manager (e.g., Blackrock), and the size of the fund. [Table A.3](#) presents our main results of [Table 2](#), controlling for these additional fund-specific variables: fund size, fund style, and fund family fixed effects.<sup>35</sup> We confirm that the 401(k) fund ownership remains economically relevant, with the coefficient hovering above 0.2, and statistically significant in any of the specifications, suggesting that our  $IO_{i,t}^{401k}$  variable is likely exogenous. Our results are also robust to adding lagged portfolio weights to the regressions, although such specification it is not justified by the demand system framework.

Lastly, we run an additional econometric test aimed at diffusing any remaining concerns. First, we estimate our original model (1) *without* our 401(k) fund-level variable. This is the same exact original specification in [Kojien and Yogo \(2019\)](#). We save the residuals  $\hat{\varepsilon}_{i,t}$  from this estimation. Next, we regress our fund-level 401(k) variable  $IO_{i,t}^{401k}$  on

---

<sup>35</sup>The *fund style* and *fund family* variables are from Morningstar.

$\hat{\varepsilon}_{i,t}$ , and save the new residuals  $\hat{\eta}_{i,t}$  from this second regressions.

By construction,  $\hat{\eta}_{i,t}$  is the component of fund-level 401(k) ownership  $IO_{i,t}^{401k}$  that is orthogonal to  $\hat{\varepsilon}_{i,t}$ , e.g., the residual from the original model. In other words, it is exogenous by construction. Lastly, we re-estimate (1) using  $\hat{\eta}_{i,t}$  (instead of  $IO_{i,t}^{401k}$ ) as regressor. The coefficient on  $\hat{\eta}_{i,t}$  is 0.232 (t-stat: 2.03), which is very similar to the one on  $IO_{i,t}^{401k}$  (0.169, t-stat: 2.39), highlighting that our fund-level 401(k) ownership variable  $IO_{i,t}^{401k}$  is likely exogenous.

<b>Panel A: All Funds</b>			
	Coefficient	s.e.	<i>t</i> -stat
$IO_{i,t}^{401k}$	0.168***	0.050	3.380
Log market-to-book	0.457***	0.058	7.830
Log book equity	0.949***	0.049	19.560
Operating profitability	0.074***	0.008	8.870
Beta	-0.064***	0.011	-5.690
Investment	-0.067***	0.010	-6.600
Dividend-to-book	-0.059***	0.011	-5.300
<b>Panel B: Mutual Funds</b>			
	Coefficient	s.e.	<i>t</i> -stat
$IO_{i,t}^{401k}$	0.240***	0.056	4.290
Log market-to-book	0.513***	0.056	9.080
Log book equity	0.810***	0.061	13.270
Operating profitability	0.033***	0.009	3.600
Beta	-0.053***	0.013	-4.190
Investment	-0.061***	0.009	-6.380
Dividend-to-book	-0.090***	0.011	-8.560
<b>Panel C: ETFs</b>			
	Coefficient	s.e.	<i>t</i> -stat
$IO_{i,t}^{401k}$	-0.060	0.055	-1.100
Log market-to-book	0.416***	0.095	4.360
Log book equity	0.961***	0.080	12.010
Operating profitability	0.067***	0.016	4.130
Beta	-0.070***	0.018	-3.840
Investment	-0.038***	0.013	-3.020
Dividend-to-book	0.007	0.021	0.340

**Table A.1: Demand system estimation - fund level  $IO_{i,t}^{401k}$ , observations not AUM-weighted.** This table reports estimates of the panel regression

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = b_0 + \beta_{0,i}\widehat{mb}_t(n) + \beta_1'X_t(n) + \beta_2 IO_{i,t}^{401k} + \alpha_t + \tilde{u}_{i,t}(n)$$

where  $\widehat{mb}_t(n)$  is the instrumented market equity-to-book equity,, and  $X_t(n)$  includes the same variables as in [Kojien and Yogo \(2019\)](#), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio.  $IO_{i,t}^{401k}$  is the 401K plans ownership of fund  $i$ , and  $\alpha_t$  are time (year) fixed effects. The funds in the regressions are not AUM-weighted. Panel A reports results using all funds, Panel B only mutual funds, and Panel C only ETFs. Variables are standardized. Standard errors are double clustered by fund and time. The sample period is from 2007 to 2020. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

<b>Panel A: All Funds</b>			
	Coefficient	s.e.	t-stat
$IO_t^{401k}(n)$	0.116***	0.007	17.060
Log market-to-book	0.580***	0.047	12.250
Log book equity	1.117***	0.041	27.140
Operating profitability	0.046***	0.006	8.030
Beta	-0.014**	0.005	-2.800
Investment	0.001	0.004	0.330
Dividend-to-book	0.008	0.005	1.640
<b>Panel B: Mutual Funds</b>			
	Coefficient	s.e.	t-stat
$IO_t^{401k}(n)$	0.122***	0.011	11.070
Log market-to-book	0.264***	0.058	4.580
Log book equity	0.751***	0.052	14.520
Operating profitability	0.066***	0.008	8.600
Beta	-0.031***	0.009	-3.560
Investment	-0.006	0.005	-1.070
Dividend-to-book	0.004	0.005	0.720
<b>Panel C: ETFs</b>			
	Coefficient	s.e.	t-stat
$IO_t^{401k}(n)$	0.082***	0.010	7.800
Log market-to-book	0.784***	0.059	13.390
Log book equity	1.326***	0.046	28.760
Operating profitability	0.034***	0.009	3.730
Beta	0.006	0.004	1.410
Investment	0.015**	0.006	2.540
Dividend-to-book	0.013	0.010	1.330

**Table A.2: Demand system estimation - stock level  $IO_t^{401k}(n)$ , observations not AUM-weighted.** This table reports estimates of the panel regression

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = b_0 + \beta_{0,i}\widehat{mb}_t(n) + \beta_1'X_t(n) + \beta_2 IO_{-i,t}^{401k}(n) + \alpha_{i,t} + \bar{u}_{i,t}(n)$$

where  $\widehat{mb}_t(n)$  is the instrumented market-to-book equity, and  $X_t(n)$  includes the same variables as in [Kojien and Yogo \(2019\)](#), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio.  $IO_{-i,t}^{401k}(n)$  is the 401K plans ownership of stock  $n$  (excluding the effect through investor  $i$ ), and  $\alpha_{i,t}$  are fund-by-year fixed effects. The funds in the regressions are not AUM-weighted. Panel A reports results using all funds, Panel B only mutual funds, and Panel C only ETFs. Variables are standardized. Standard errors are double clustered by fund and time. The sample period is from 2007 to 2020. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

	(1)	(2)	(3)
$IO_{i,t}^{401k}$	0.23*** (3.42)	0.26*** (3.29)	0.18** (2.33)
Log market-to-book	0.75*** (7.89)	0.70*** (6.87)	0.64*** (6.25)
Log book equity	1.27*** (15.29)	1.21*** (13.32)	1.18*** (13.59)
Operating Profitability	0.03** (2.66)	0.04** (2.84)	0.05*** (4.09)
Beta	-0.03** (-2.87)	-0.03** (-2.60)	-0.03** (-2.83)
Investment	-0.01 (-1.07)	-0.01 (-1.20)	-0.01 (-0.74)
Dividend-to-book	-0.04** (-2.96)	-0.04*** (-3.01)	-0.02 (-1.59)
Fund size	-0.40*** (-5.87)	-0.30*** (-4.24)	-0.22*** (-3.72)
Time fixed effect	Yes	Yes	Yes
Fund family fixed effect	No	Yes	Yes
Fund style effect	No	No	Yes

**Table A.3: Robustness for fund level  $IO_{i,t}^{401k}$ .** This table reports estimates of the panel regression

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = b_0 + \beta_{0,i}\widehat{mb}_i(n) + \beta_1'X_t(n) + \beta_2 IO_{i,t}^{401k} + Fundsize_{i,t} + \alpha_t + \bar{u}_{i,t}(n)$$

where  $\widehat{mb}_i(n)$  is the instrumented log market equity-to-book equity, and  $X_t(n)$  includes the same variables as in [Kojen and Yogo \(2019\)](#), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio.  $IO_{i,t}^{401k}$  is the 401K plans ownership of fund  $i$ , and  $\alpha_t$  are time (year) fixed effects. We include additional fund-level controls to rule out potential endogeneity concerns.  $Fundsized_{i,t}$  is the log value of a fund's assets under management. Column (1) reports results with time fixed effects, while column (2) reports results with time fixed effects and fund family fixed effects. Column (3) also includes fund style fixed effects. The funds in the regressions are AUM-weighted. The sample period is from 2007 to 2020. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

# Internet Appendix B Holdings Data: Additional Analysis and Robustness

## B.1 Thomson Reuters s34 Holdings

The analysis in [Section 4.1](#) and [Section 4.2](#) relies on data from Morningstar, which provides detailed holdings of *individual* mutual funds and ETFs. Instead, the analysis in [Section 4.4](#) relies on the Thomson Reuters' s34 file, which provides aggregated holdings of *all funds* under the manager's control ([Kojien and Yogo \(2019\)](#), [Kojien et al. \(2022\)](#)).

[Table B.1](#) repeats the same analysis presented in [Table 4](#) but at the fund family level, i.e., using the s34 data. Importantly, the coefficient on stock-level ownership remains large and significant. In particular, we find that the coefficient of .218 is close in magnitude to the one reported in Panel (b) of [Table 4](#), despite the fact that s34 fund-family holdings blend together ETFs and mutual funds.

Thomson Reuters (s34) holdings			
	Coefficient	s.e.	t-stat
$IO_t^{401k}(n)$	0.218***	0.037	5.830
Log market-to-book	1.514***	0.162	9.350
Log book equity	1.893***	0.067	28.300
Operating profitability	0.006	0.007	0.830
Beta	0.056*	0.030	1.840
Investment	0.036*	0.020	1.810
Dividend-to-book	-0.135***	0.031	-4.300

**Table B.1: Demand system estimation - Stock level  $IO_t^{401k}(n)$  with s34 holdings.** This table reports estimates of the panel regression

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = b_0 + \beta_{0,i}\widehat{mb}_t(n) + \beta_1'X_t(n) + \beta_2 IO_t^{401k}(n) + \alpha_{i,t} + \bar{u}_{i,t}(n)$$

where  $\widehat{mb}_t(n)$  is the instrumented market-to-book equity, and  $X_t(n)$  includes the same variables as in [Kojien and Yogo \(2019\)](#), e.g., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio.  $IO_t^{401k}(n)$  is the 401K plans ownership of stock  $n$ , and  $\alpha_{i,t}$  are manager-by-year fixed effects. The mutual fund institutions in the regressions are AUM-weighted. Variables are standardized. Standard errors are double clustered by fund institution and time. The sample period is from 2007 to 2020. \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

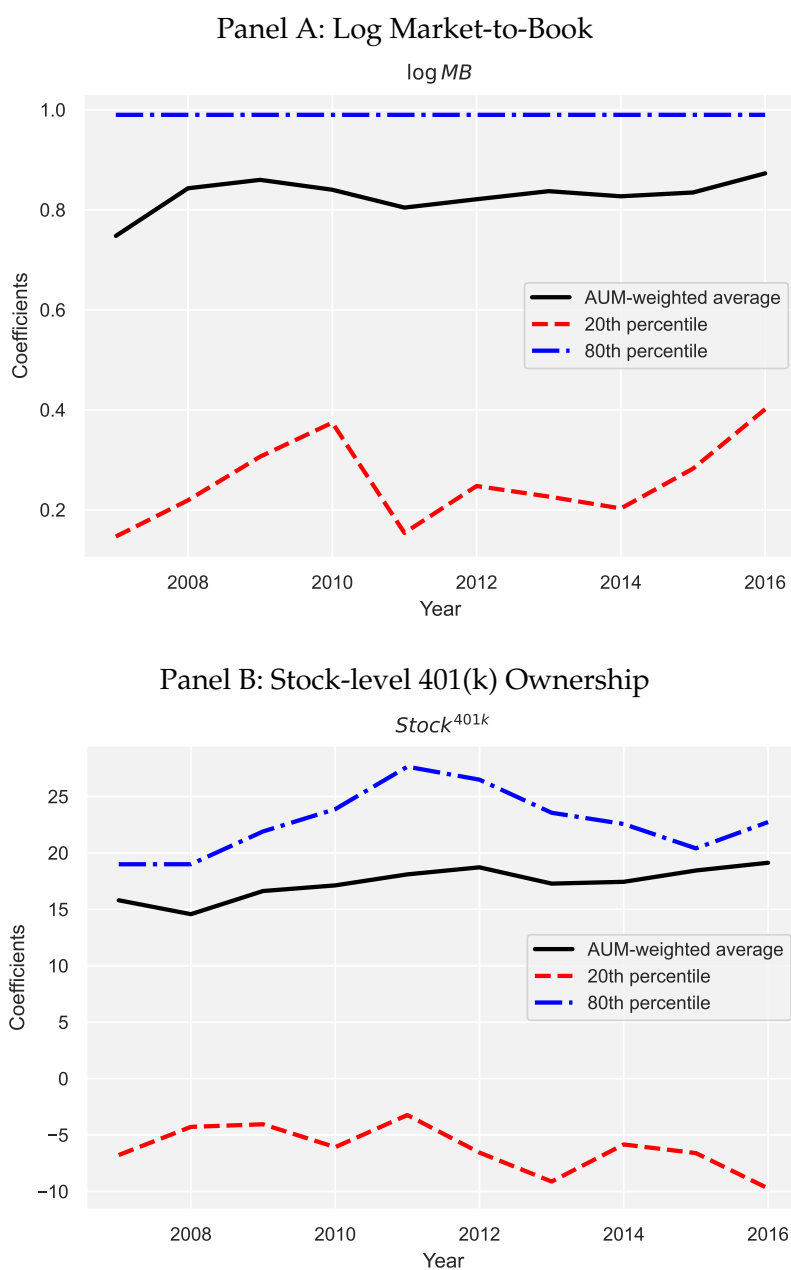


The coefficients on the other characteristics, e.g., beta, investment, and dividend-to-book are small in both datasets. Overall, it appears that the empirical results using the s34 dataset are in line with those reported in [Section 4.2](#) and, thus, the analysis in [Section 4.4](#) is informative of the equilibrium price impact of a change in 401(k) stock-level ownership.

## B.2 Holdings scraped directly from 13F filings

In this section, we repeat our computation of the equilibrium price impact presented in [Section 4.4](#) using the 13F filings data provided by [Backus et al. \(2021\)](#). These authors collected 13F filings from the SEC's EDGAR database since electronic filing was made mandatory in 1999, and addressed gaps in coverage and errors that appear in commercial datasets of institutional holdings (e.g., Thomson Reuters). The disadvantage of such dataset is that we cannot anymore exploit the [Kojien and Yogo \(2019\)](#) classification of institutions into six types. Thus, in the estimation, we abstract from investor types and (1) keep institutions with more than 1,000 strictly positive holdings separate; (2) group institutions with fewer than 1,000 holdings based on TNA, so that each group has on average 2,000 holdings.

[Figure B.1](#) is the counterpart of [Figure 5](#). Importantly, both the coefficient governing the elasticity of demand, and the coefficient on 401(k) stock-level ownership display a similar range in terms of magnitude across the two datasets. It is therefore not surprising that the cross-sectional distribution of aggregate price impact across stocks reported in the top left panel of [Figure B.2](#) remains economically sizable: a one standard deviation increase in 401(k) ownership, around 1.3% in 2007 and 1.6% in 2016, leads to a price impact (for the median stock) slightly less than 40 percent in 2007 and about 90 percent in 2016. Similarly to the s34 dataset, we observe a stronger price impact for large stocks (bottom right panel), with a sharp increase in 2015, and little difference for stocks sorted on market betas (bottom left panel). The main difference across the two datasets is observed for stocks sorted on book-to-market. In particular, the scraped data of [Backus et al. \(2021\)](#) suggest a larger impact for growth stocks.



**Figure B.1: Price impact: relevant coefficients.** This figure shows the annual coefficient on log market-to-book (top panel) and stock-level 401(k) ownership (bottom panel) for financial institutions in [Backus, Conlon and Sinkinson \(2021\)](#), estimated annually, by GMM with zero weights. Variables are standardized (within each year) to make coefficients comparable. We report the cross-sectional mean of the estimated coefficients by institutions, weighted by assets under management. The coefficient on 401(k) ownership is multiplied by 100. The sample period is from 2007 through 2016.



**Figure B.2: Institutional price impact.** This figure shows the price impact of a change in the stock-level 401(k) ownership estimated through the diagonal elements of the matrix  $M$  defined in (9) using holdings data from [Backus, Conlon and Sinkinson \(2021\)](#). The top left panel shows the 10th percentile, median and 90th percentile of price impact across all stocks. The remaining panels plot the average price impact across the stocks in the top quintile and bottom quintile of portfolios sorted on (top right), beta (bottom left), and size (bottom right), using NYSE break points as cutoffs.

## Internet Appendix C Data Cleaning Procedure

### C.1 BrightScope and Morningstar (MS)

1. We match funds held in 401(k) plans with Morningstar holdings.
2. We remove mutual funds whose portfolio weights as reported by Morningstar are different from the correct portfolio weights calculated using holdings values and total net assets, as in [Pástor et al. \(2015\)](#).
3. We merge fund characteristics (e.g., fund TNA) from Morningstar with the dollar allocation of 401(k) plans to funds from BrightScope. We then calculate our  $IO_{i,t}^{401k}$  variable, a fund's 401(k) ownership. We drop funds where  $IO_{i,t}^{401k} < 0$  or  $IO_{i,t}^{401k} > 1$ .
4. Our analysis focuses on equities, hence we only keep equity mutual funds having an equity ratio  $\geq 0.75$ .
5. We merge fund holdings with firm data from CRSP and COMPUSTAT, replacing missing dividends as zero.
6. We drop fund-stock observations with missing characteristics.
7. We define the investment universe for each fund as described in the paper. We only keep funds with clearly defined investment universes (e.g., the number of stocks in the investment universe is greater than zero)
8. We drop funds holding fewer stocks than the fifth percentile in the cross-section of funds, every year (approx. 15 stocks).
9. As in [Kojen and Yogo \(2019\)](#), each year, we winsorize profitability, investment, and market beta at the 2.5th and 97.5th percentiles to reduce the impact of outliers. Since dividends are positive, we winsorize dividends to book equity at the 97.5th percentile. We also winsorize  $\log(\text{book equity})$  at the 2.5th and 97.5th percentiles.
10. We winsorize funds' total net assets (TNA) at the 97.5th percentile, every year, to deal with outliers.

11. In the GMM estimation, we keep zero-weight holdings, e.g., stocks in a fund's investment universe, but currently not being held by the fund. Zero-weight holdings must have non-missing characteristics.

### Estimation

- In the pooled regressions, we implement 2SLS with instrumented log market-to-book, and use fund TNA as weights.
- For GMM, we include zero holdings of a stock, and use fund TNA as weights.
- As in [Kojien et al. \(2022\)](#), we impose the economic constraint  $\log(\text{MB}) < 1$  in all the estimations.
- The price impact analysis is based on yearly GMM estimations.

## C.2 Thomson Reuters s34 Holdings

1. We use the same institutional types as in [Kojien and Yogo \(2019\)](#).
2. We merge the s34 holdings data with CRSP and COMPUSTAT.
3. We define the investment universe for each institution.
4. In the GMM estimation of price impact, we pool institutions into groups by type and TNA as in [Kojien and Yogo \(2019\)](#), include holdings with zero weights (e.g., belonging to the investor's investment universe, but not currently owned), and calculate the instrument based on these pooled groups.

## C.3 Scraped Holdings from 13F Filings

1. We follow [Backus et al. \(2021\)](#), and use their 13F scraped holdings between 2007 and 2016.
2. We merge these holdings with CRSP and COMPUSTAT, and define the investment universe for each institution.

3. We drop institutions holdings less than 100 stocks at any given time, and pool institutions into groups by TNA as in [Haddad et al. \(2022\)](#). We then calculate the instrument based on these pooled groups.
4. We estimate the price impact via GMM, including holdings with zero weights (e.g., belonging to the investor's investment universe, but not currently owned).