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

Riccardo Sabbatucci†
Stockholm School of Economics

Andrea Tamoni‡
Rutgers Business School

Song Xiao§
LSE

August 28, 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) stock holdings on investor behavior. We introduce a new variable, *stock-level 401(k) ownership*, and find it to be a key determinant of investor demand, with a one standard deviation increase in 401(k) ownership leading to 11-19% increase in stock demand. We also estimate the equilibrium price impact of a change in stock-level 401(k) ownership to be positive and increasing over time, consistent with the shift from active to passive investing. Lastly, we document that funds managing a larger fraction of 401(k) assets tilt their portfolios toward winners, high beta and long duration stocks, and they hold less cash.

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 Herbert Business School (University of Miami), and Rutgers Business School.

†Department of Finance, Stockholm School of Economics and Swedish House of Finance (SHoF). **Email:** riccardo.sabbatucci@hhs.se

‡Department of Finance, Rutgers Business School. **Email:** atamoni@business.rutgers.edu

§PhD Student, Department of Finance, London School of Economics. **Email:** 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 adopting a demand-based framework, following the seminal work of Koijen and Yogo (2019). Demand-based explanations linked to the role of investors' heterogeneity have been recently proposed to explain variation in domestic equities (Koijen and Yogo, 2019), corporate bonds (Bretscher et al., 2021), and exchange rates (Koijen and Yogo, 2020), and to analyze the shift from active to passive investing (Haddad et al., 2022), green investments (Koijen et al., 2022), and the portfolios of high-net worth individuals (Gabaix, Koijen, Mainardi, Oh and Yogo, 2022). A key element 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 light on this important component. More precisely, using a demand-based framework, we study the impact that 401(k) pension plans have on investors' demand for stocks, and introduce a novel variable that appears to be a key determinant of investors' allocation decisions: *stock-level 401(k) ownership*. We then present two channels through which 401(k) ownership might drive investors' demand for specific stocks, and test their relevance.

The first channel through which 401(k) plans could affect the demand of individual stocks is related to the magnitude of stock-level 401(k) ownership. The fraction of an individual stock cumulatively owned by 401(k) plans can be seen as an additional stock characteristic, similarly to book-to-market or momentum. Fund managers and other types of investors might take into account the information conveyed by this additional stock characteristic when evaluating how many shares of a specific stock to purchase. For example, it might be that active funds prefer to deviate from their respective benchmarks by investing in stocks with more stable and long-term investors, like pension funds. Thus, we hypothesize that the quantity of a stock cumulatively owned by 401(k) plans might affect the demand for that specific stock. We label this the *stock level* channel.

The second channel through which 401(k) allocations could impact individual stocks' demand is by direct flows to mutual funds and ETFs, which, in turn, use that cash to increase their equity exposure. We hypothesize that funds managing 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, and hold less cash. For example, some funds display a preference for stocks with high market beta (e.g., Christoffersen and Simutin (2017), Han et al. (2022)). We label this the *fund level* channel.

We report several interesting findings. First, 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 size – in explaining the demand of mutual funds and ETFs for a specific stock. In response to a one standard deviation increase in 401(k) stock ownership, the average active mutual fund demands approximately 18.6% (*t*-stat: 11.16) more of the stock. The average active ETF also increases exposure to the stock by 11.5% (*t*-stat: 10.21) for each standard deviation increase in 401(k) stock ownership. Importantly, stock-level 401(k) ownership appears to be *distinct* from other forms of institutional ownership, such as total mutual fund (Chen et al., 2000) or largest (top 10) investors' ownership of a stock (Ben-David et al., 2021). In fact, after controlling for these alternative types of ownership, the magnitude of the coefficient on our stock-level 401(k) ownership is barely affected, and so is its statistical significance. These results highlight the unique information content of stock-level 401(k) ownership for fund managers' decision. To the best of our knowledge, this is the first paper to highlight the specific role of stock-level pension allocations for fund manager decisions.

Motivated by the importance of stock-level 401(k) ownership for funds' investment decisions, we then explore the equilibrium price impact of a change in stock-level 401(k) ownership for the cross-section of stocks, over time. We estimate 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 over the last decade documented in the literature (Koijen et al., 2022). To further validate the direct impact 401(k) ownership has on individual stocks, we employ a matching procedure to pair stocks with similar fundamental characteristics but different levels (e.g., high vs. low) of 401(k) ownership. We find that stocks with positive 401(k) ownership tend to earn 3%-5% higher annual returns than similar stocks, in terms of characteristics and investor structure, not owned by 401(k) plans.

We also highlight the importance of stock-level 401(k) ownership by studying the impact that trading by 401(k) plans has on individual stock returns using two additional tests. Our first methodology exploits large changes in individual stock holdings by aggregate 401(k) plans (Ben-David et al., 2021). We find that large changes in the holdings of an individual stock substantially affect its contemporaneous return. A large positive position (trade) taken by 401(k) plans in a stock generates a 12% higher return than that implied by an average trade, followed by a partial return reversal over the next two years. Our second test exploits the granular instrumental variable approach introduced by Gabaix and Koijen (2022). We find that a 10% increase in the (instrumented) stock demand of 401(k) plans generates an average stock price increase of 3.6%, after controlling for standard firm-specific drivers of stock returns.

Second, we analyze the fund-level channel of 401(k) ownership, and document that funds managing a larger fraction of 401(k) assets display greater demand for stocks. More in detail, 401(k) fund ownership is the most important variable, after firm size, in explaining how much of a stock funds demand. For one standard deviation increase in 401(k) fund ownership, the average active 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.

To gain further insight on the relation between 401(k) assets and fund managers' portfolio allocations, we then study how the investment strategies and performance of active 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 could, for example, 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 the alpha (statistically) unaffected. 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 test whether investment funds perceive 401(k) flows to be more stable by studying the level of their cash holdings. We find this to be indeed the case. Mutual funds managing a large fraction of 401(k) assets have approximately 32% *less* cash holdings compared to other mutual funds.

Our paper is related to the emerging demand-based asset pricing literature. Koijen 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. Focusing on mutual funds, Ben-David et al. (2022b) show that their ratings generate correlated demand that creates systematic price fluctuations. Furthermore, exploiting a reform to the Morningstar rating system, Ben-David et al. (2022a) also document that demand effects generated by institutional frictions can influence systematic return predictability patterns in stocks and mutual funds.

Our contribution to this strand of the literature is to highlight the unique relevance of stock-level and fund-level 401(k) ownership 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. 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 a key component of fund size is determined by 401(k) plans allocations, 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. In contrast to their paper, we document that a large fraction of pension assets managed by a fund influences its performance 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. Pool et al. (2016) study whether mutual fund families acting as service providers in 401(k) plans display favoritism toward their own affiliated funds, and find that fund deletions and additions are less sensitive to prior performance for affiliated than unaffiliated funds. Differently from both studies, we do not focus on investment menu offered by plans, but directly estimate the demand of individual stocks by funds offered in 401(k) menus using a demand-based framework. Our quantification of the price and return impact of the stock-level channel of 401(k) ownership constitutes a unique contribution to this literature.

Moreover, whereas Sialm et al. (2015) and Christoffersen and Simutin (2017) rely on survey data about DC assets from "Pensions & Investments" (P&I) administered to domestic equity funds, we instead 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 and Section 5 present the stock and fund level results, respectively. Section 6 concludes.

2 Data

2.1 Data Sources

Our 401(k) plan holdings data comes from BrightScope Beacon. BrightScope Beacon provides comprehensive plan-level holdings data gathered from audited Form 5500 filings of private-sector defined contribution (DC) plans from 2007 to 2020. We focus exclusively on 401(k) defined contribution plans in this paper.[1] BrightScope reports annual data on

[1]BrightScope Beacon also provides holdings for 403(b) plans, although their total market value is small relative to that of 401(k) plans.

the investment options (e.g., mutual funds) available to plan participants together 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 2007-2020, 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.[2] We match mutual funds in 401(k) plans with Morningstar by fund tickers and names.[3]

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.[4] We also require funds to have at least 3 years of holdings data. We include only equity mutual funds and ETFs directly owned by 401(k) plans.[5] Our final dataset comprises a total of 2,156 funds, split between 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 Koijen 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 Koijen et al. (2022). Profitability is defined as operating profits scaled by book value of

[2]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.

[3]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.

[4]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).

[5]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.

equity, investment as the annual growth rate of total assets, and dividends-to-book equity 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),[6] 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.[7] 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 financial companies to DC plan sponsors.

The mutual fund category, which also includes ETFs, is the largest component comprising, on average, 43% of the total 401(k) assets. Figure 2 decomposes this 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

[6] 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.

[7] 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.

together in the "Others" category. US index funds include mutual funds and ETFs that are index-tracking.[8]

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 fund assets are owned, on average, by 401(k) plans, making 401(k) plans among the largest fund investors. The second row shows that 401(k) plans represent large investors also when we focus onto the universe of active mutual funds. 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 fifth 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 fund-level 401(k) ownership, $IO^{401k}_{i,t}$, over time. We notice that the variable is stationary, even at the 75th percentile of the distribution, where it fluctuates between 6% and 12%.

[8]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).

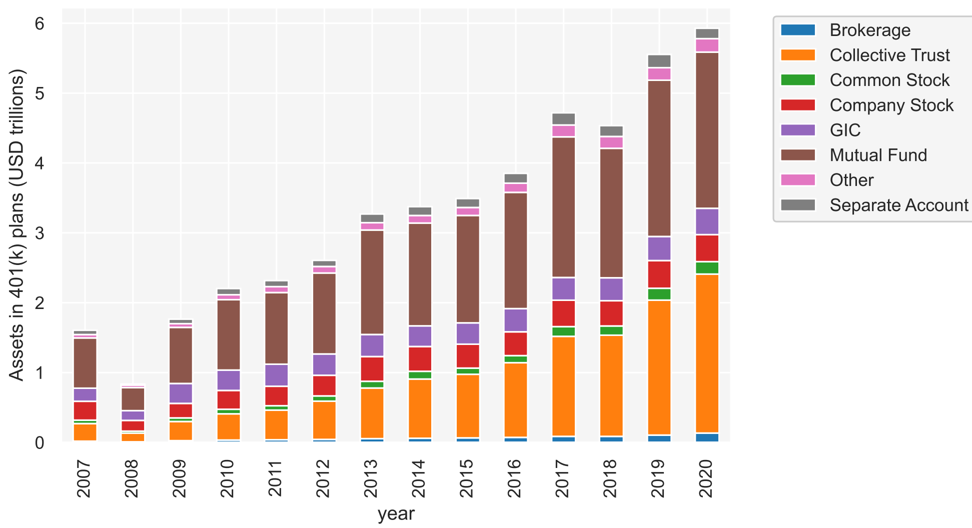


Figure 1: 401(k) plan assets. This figure shows the distribution of 401(k) plan assets into the various investment options, over time. Annual data, from 2007 to 2020.

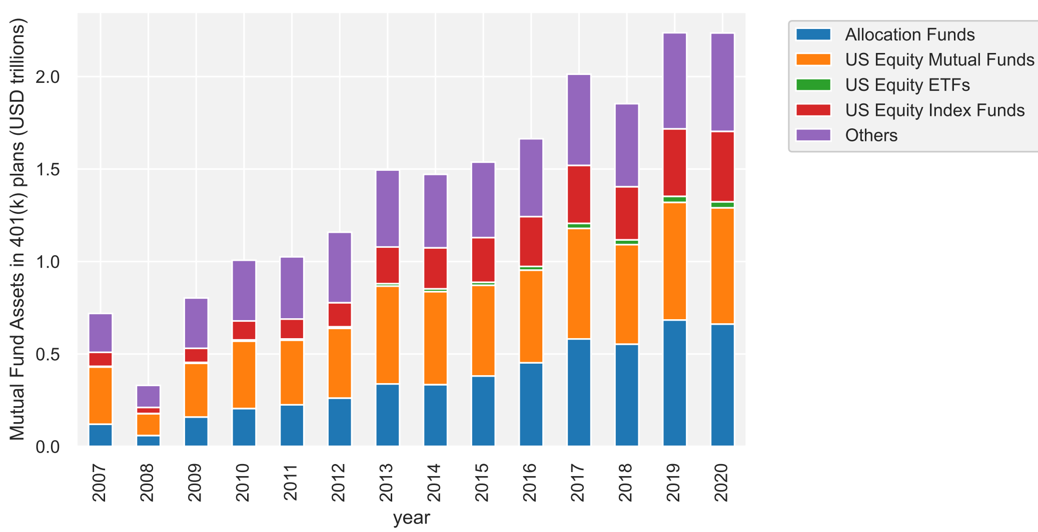


Figure 2: Distribution of assets within the mutual fund category. This figure plots the value of 401(k) mutual fund investments split into various subgroups. Allocation funds are balanced funds investing in a mix of fixed income assets and equities depending on their objective, e.g., target-date funds. 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.

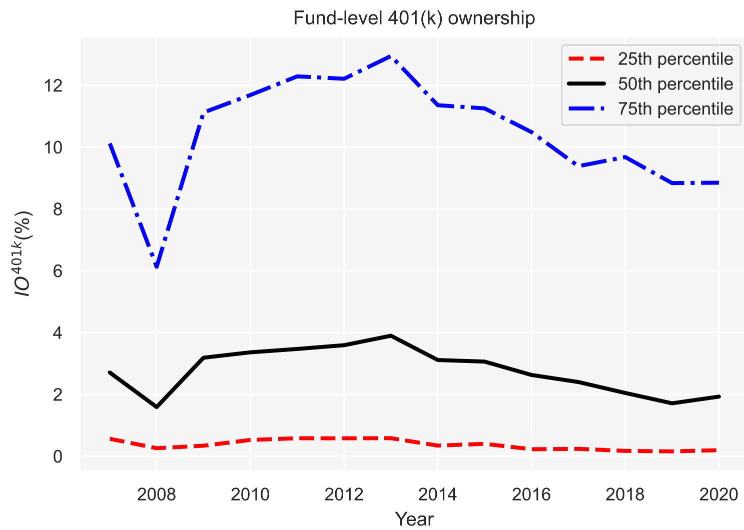


Figure 3: Fund-level 401(k) ownership over time. This figure shows the cross-sectional distribution of fund-level 401(k) ownership over time. Annual data, from 2007 to 2020.

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

As discussed in Section 2, 401(k) plans invest a substantial amount of assets in equity mutual funds and ETFs. In this section, we outline how we adapt the asset demand framework of Koijen 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 Koijen and Yogo (2019), and define the demand curve of investor i for stock n as:

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

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 k observed characteristics of asset n at date t, and $w_{i,t}(0)$ is the portfolio weight on the outside asset. As in Koijen 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 Koijen and Yogo (2019) with one additional variable, potentially capturing variation in investor demand: 401(k) ownership of either the individual stock n (denoted $IO_t^{401k}(n)$), or the investor i (denoted $IO_{i,t}^{401k}$).

Following Koijen and Yogo (2019), we assume throughout that stock characteristics are exogenous to latent demand,

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

By explicitly controlling for variables such as book-to-market and profitability in the regressions, and instrumenting market equity as in Koijen 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.

Nevertheless, latent investor demand is likely correlated with a stock's market capitalization, i.e., $\mathbb{E}_t\left[\epsilon_{i,t}(n) \mid me_t(n)\right] \neq 0$, because some investors are large and their individual latent demand affects stock prices.[9] Hence, the model in (1) delivers biased and inconsistent estimates.

We therefore construct an instrument $z_{i,t}(n)$ for the endogenous market capitalization. Specifically, we follow Koijen 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, $\mathcal{N}_{i,t} = \cup_{k=0}^{2}\mathcal{S}_{i,t-k}$ where k is expressed in years.[10]

[9] Market equity is the numerator of log market-to-book equity.

[10] For each fund i, the outside asset includes the complement set of stocks, those not in the investment universe.

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 investorâĂŹs own holdings:[11]

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

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 investorâĂŹs choice set.[12]

3.2 Economic Channels

The first channel through which 401(k) plans affect the demand of individual stocks is by ownership of the stock itself.[13] We can think of the fraction of a stock held by 401(k) plans as a 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 company to purchase.[14] In other words, the total percentage ownership of a stock by 401(k) plans might affect fund demand for that specific stock. We study this channel, labeled the *stock level* channel, in Section 4.

The second channel through which 401(k) allocations affect individual stock demand

[11]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.

[12]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.

[13]We only focus on indirect ownership, e.g., through funds, since direct stock ownership by 401(k) plans of individual stocks is usually negligible.

[14]Institutional ownership of a stock is a characteristic known amongst investors; in particular, 401(k) plans ownership can be retrieved by public filings or third party data providers.

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 such as hedge funds, mutual funds and ETFs managing a larger fraction of 401(k) assets might have a more stable investor base.[15] As a consequence, they might invest in different types of stocks compared to funds managing with fewer 401(k) assets, all else being equal. In other words, the "investor base" of mutual funds and 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. We study this channel, labeled the *fund level* channel, in Section 5.

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 stock n owned in aggregate by 401(k) plans, and the fraction of fund i's assets under management owned by all 401(k) plans at time t. Next, we estimate the magnitude of these demand effects.

4 Stock Level Channel

We start our analysis by studying the relevance of 401(k) ownership at the individual stock level. This firm-specific characteristic, similarly to, for example, book-to-market or beta, might help explain how much of a stock is demanded by funds. 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,[16] and individual fund allocations do not drastically change over time,[17] then a larger stock ownership by 401(k) plans could signal potential stability in the

[15] 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 greater than 0.8.

[16] 401(k) plans have persistent fund allocation in terms of proportions of assets under management, see Section 2.

[17] 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

stock investors' base as well. To this end, we calculate the fraction of stock n cumulatively owned by 401(k) plans:

$$IO_t^{401k}(n) = \frac{\sum_{j=1}^{I} IO_{j,t}^{401k} * w_{j,t}(n) * AUM_{j,t}}{ME_t(n)} \tag{3}$$

where $IO_{j,t}^{401k}$ is the fraction of fund j owned by all 401(k) at the end of year t, $w_{j,t}(n)$ 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:

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

where 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 a vector of controls that includes the firm-specific characteristics specified in Koijen and Yogo (2019), and $IO_{-i,t}^{401k}(n)$ is the fraction of stock n cumulatively owned by 401(k) plans through funds, excluding that owned by fund i. Note that by excluding investor i from the $IO_t^{401k}(n)$ regressor in (3), we are studying how the portfolio choice of fund i is influenced by the stock-level 401(k) ownership through all *other* investors, thus reducing possible endogeneity concerns.[18] In addition to the variables used in Koijen and Yogo (2019), we also report results controlling for two additional stock characteristics that potentially impact fund demand for individual stocks: the fraction of a stock owned by the top ten investors (Ben-David et al., 2021), and a stock's total mutual fund ownership.

Table 2 shows the results from the panel regression (4) for the entire universe of funds (columns (1)-(3)), mutual funds (columns (4)-(6)), and ETFs (columns (7)-(9)). We report two-way (funds and time) clustered standard errors. Fund-stock observations are

---

also held in the previous quarter.

[18]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.

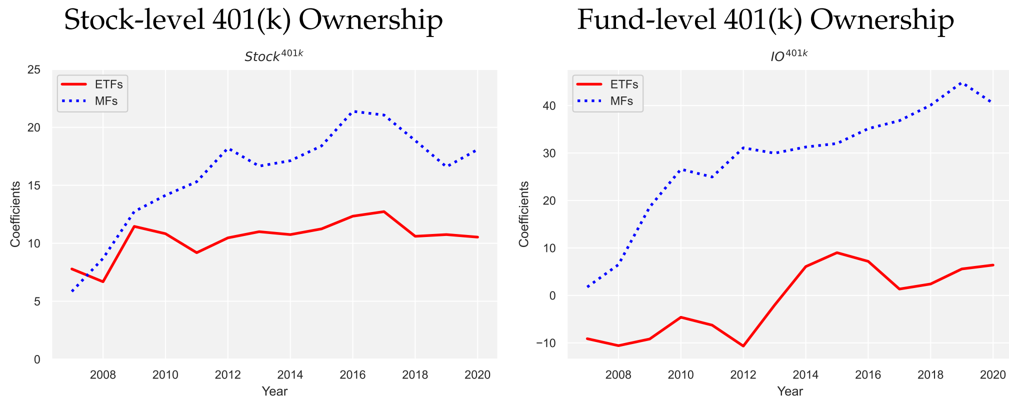


Figure 4: Coefficients on 401(k) ownership. This figure shows the annual coefficient in equations (4) (stock level, left) and (10) (fund level, right) and on 401(k) ownership, separately for mutual funds and ETFs, estimated by pooled OLS using assets under management as weights. The regression is estimated annually, and it includes fund-level fixed effect in the left 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.

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 after size (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-stat=18.95, column (1)) 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.[19] Controlling for stock ownership by the top ten investors, or total mutual fund ownership of a stock, does not affect our results, highlighting the uniqueness and relevance of 401(k) stock-ownership with respect to other types of institutional ownership.[20]

The left panel in Figure 4 shows the evolution of the coefficient on stock-level 401(k)

[19] Table A.1 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.

[20] In Internet Appendix B.1 we verify that our stock-level 401(k) results are robust to using s34 data instead of Morningstar.

ownership, $IO^{401k}_{-i,t}(n)$, over time.[21] 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 2. Panel A of Table 3 reports estimates from the GMM estimation of the non-linear version of equation (4). The result shows a positive on stock-level 401(k) ownership of 0.25 – the second largest within the set of characteristics $\mathbf{X}_t(n)$ – statistically significant at the 1% level (t-stat: 25.49).

Next, we study the equilibrium price impact of stock-level 401(k) ownership.

4.1 Equilibrium 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

$$\frac{\partial p_t(n)}{\partial IO^{401k}_t(n)} \tag{5}$$

where p is the log price of stock n. Following Koijen 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 $\mathbf{M}$:[22]

$$\mathbf{M} = \left(\mathbf{I} - \sum_i \beta_{0,i} A_i \mathbf{H}^{-1}\mathbf{G_i}\right)^{-1} \left(\sum_i \beta_{2,i} A_i \mathbf{H}^{-1}\mathbf{G_i}\right), \tag{6}$$

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 assets 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 (6) is the aggregate demand elasticity (Koijen and Yogo, 2019), and its diagonal elements are strictly positive when $\beta_{0,i} < 1$ for all investors. If a firm is held by

[21] Figure A.1 shows the coefficients on the other covariates.

[22] To compute this expression one has to exploit the identity $\mathbf{p} = \log\left(\sum_i A_i \mathbf{w}_i\right) - \mathbf{s}$ (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.

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 ownerhship (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) ownerhship is a weighted average of $IO_t^{401k}(n)$ coefficients of its institutional owners, adjusted for their demand 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 Koijen and Yogo (2019) classification of institutions into six types (i.e., $i = 1, \ldots, 6$): 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, since it reports aggregate holdings at the investor level (e.g., for all funds managed by Fidelity).[23]

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 by GMM, accounting for zero holdings, under moment condition (2).[24] We confirm the results in Koijen and Yogo (2019) that mutual funds have less elastic demand than investment advisors for most of our 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

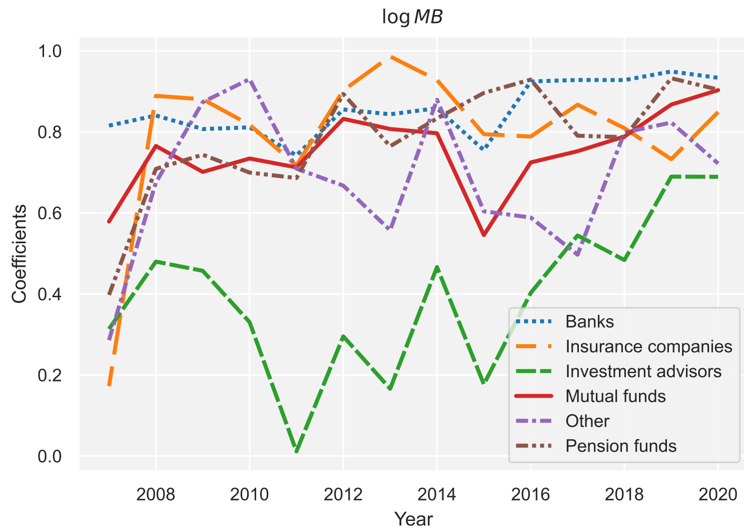
[23]Mindful of potential gaps in coverage of institutional holdings in the s34 files, we validate our price impact results by replacing s34 data with data on 13F filings from Backus et al. (2021) in Internet Appendix B.2.

[24]To obtain the price impact, we estimate the coefficients as in Koijen and Yogo (2019). For institutions with more than 1,000 stocks in their holdings, we estimate coefficients by institution. For the remaining institutions, we group them by type (e.g., mutual funds) such that on average each group holds 2,000 stocks at any point in time. Variables are standardized within each institution (or group) and for each year.

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). 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 endogeneity concerns: we use stock holdings of banks, insurance, etc. as left hand side variables, while we employ only mutual funds and ETFs holdings in the construction of our stock-level 401k ownership (right hand side variable).

Given estimates of $\beta_{0,i,t}$ and $\beta_{2,i,t}$ for each investor, we can calculate, each 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 all 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 sample period. The stronger effect over time can be related to the shift from active to passive investing of the last decade, since equation (6) 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

Panel A: Log Market-to-Book



Panel B: Stock-level 401(k) Ownership

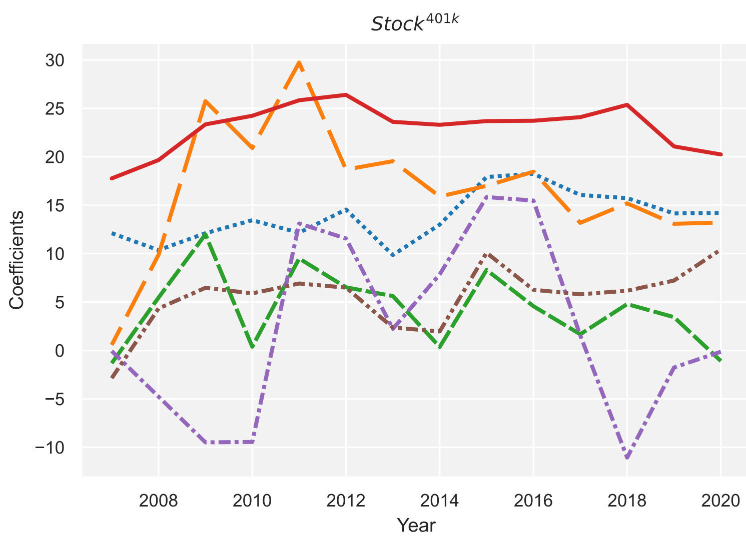


Figure 5: Price impact: relevant coefficients. This figure shows the annual coefficient on log market-to-book (top panel) and stock-level 401(k) ownership (bottom panel) for financial institutions in Thomson Reuters holding (s34) estimated annually by GMM with zero weights. Variables are standardized (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.

than for small stocks), given tracking error concerns.[25]

[25] In the U.S. stock market, large corporations like Apple make up a substantial fraction of total market capitalization and, as a consequence, a large change in those portfolio weights would cause a substantial impact on an institution's total portfolio return.

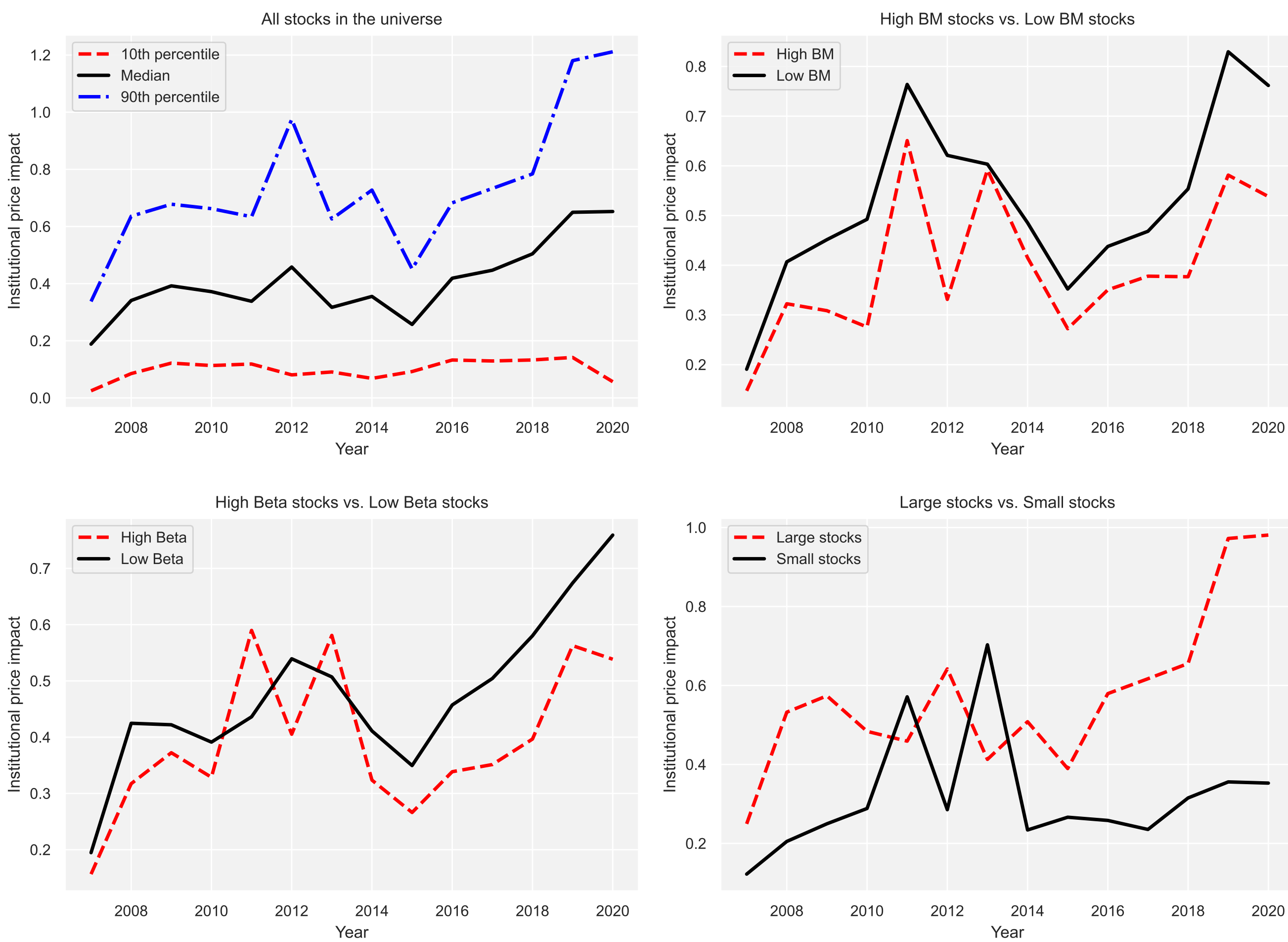


Figure 6: Institutional price impact. This figure shows the price impact of a change in the stock-level 401(k) ownership estimated through the diagonal elements of the matrix **M** defined in (6). The top left panel shows the 10th percentile, median and 90th percentile of price impact across all stocks. 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.

We do not observe noticeable differences for stocks sorted on book-to-market or betas, which suggests that our 401(k) ownership variable 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 this trend, particularly for value and high-beta stocks.

4.2 Matched Sample of Low and High Stock-level 401(k) Ownership

In Section 4, we estimated the impact of 401(k) plans for individual stock demand using the framework of Koijen and Yogo (2019). In this section, instead, we quantify the direct impact 401(k) ownership has on individual stock *returns* by employing a matching analysis: we compare otherwise identical stocks that only differ by 401(k) ownership, and analyze their return dynamics. In other words, we match pairs of similar stocks together, one displaying positive 401(k) ownership (the treated stock), while the other not owned by 401(k) plans (the control stock). This matching exercise allow us to evaluate whether stocks belonging to the treatment and control groups, which are otherwise identical, perform differently.

We start our analysis identifying, every year, the largest institutional investor for each stock (e.g., Blackrock, Fidelity, etc.).[26] For each stock, we end up with a time series of its largest institutional investor. We then count the number of times each investor is ranked as the top one across all stocks and years, and extract the top ten investors' names. This list includes Blackrock, Vanguard, Fidelity, Dimensional Fund Advisors, among others.

Next, for each of these ten investors, we select the subset of stocks for which this investor (e.g., Vanguard) is the largest. Within this subset of stocks, we match stocks with positive 401(k) ownership (*treated* stock) with a comparable group made of stocks without 401(k) ownership (*control* stock). Comparable stocks share the same largest investor (e.g., Vanguard), and have similar (i) portfolio weights in the largest investor's portfolio; (ii) size; (iii) book-to-market. More precisely, we sort the candidate "matching" stocks

[26]Since 13F holdings are quarterly, we select the investor ranked at the top most of the quarters within a year. If there is a draw, we select the largest investor in terms of AUM.

on the difference between their market capitalization and the treated stockâĂŹs market capitalization. This generates a "market cap rank," where the candidate stock with rank = 1 has a market cap closest to the one of the treated stock. We repeat the same ranking methodology with respect to the book-to-market. We then select the stock with the *smallest* sum of market cap and book-to-market ranks for each treated stock, every year, and include the "matched" stock in our control group.

We repeat the above matching procedure for each stock owned by all of the ten largest investors. Lastly, we estimate a panel regression of annual returns of the matched pair of stocks on a treated dummy variable and various stock-level controls. Controls include lagged size, book-to-market, beta, and momentum. Standard errors are double clustered by stock and time.

Panel A of Table 4 report the average characteristics of matched sample, while Panel B the regression results. The coefficient on the "treated dummy" is slightly above 5% in a specification without controls, and around 3.2% after controlling for main drivers of cross-sectional return variation, such as beta, book-to-market, log market equity and momentum. In other words, stocks with positive 401(k) ownership tend to earn 3%-5% more than similar stocks – in terms of characteristics and investor structure – not owned by 401(k) plans.

4.3 Price Impact of Trades by 401(k) Plans

Next, we study the impact of trading by 401(k) plans on individual stock returns. Similarly to Ben-David et al. (2021), for each stock-year pair, we calculate the percentage change in shares held by 401(k) plans. Then, we construct a large (small) trade dummy for 401(k) plans if the stock is in the top (bottom) quintile of the cross-sectional distribution of 401(k) trades for the year. We repeat the same exercise for the cumulative ownership of the top ten investors in every stock. By focusing on large changes in holdings of a stock by 401(k) plans, we identify positions that are actively traded, where the price impact of pension plans might be more relevant.

We then run a regression of individual annual stock returns on actively and non-actively traded 401(k) and top10 investors dummies, controlling for log size and time

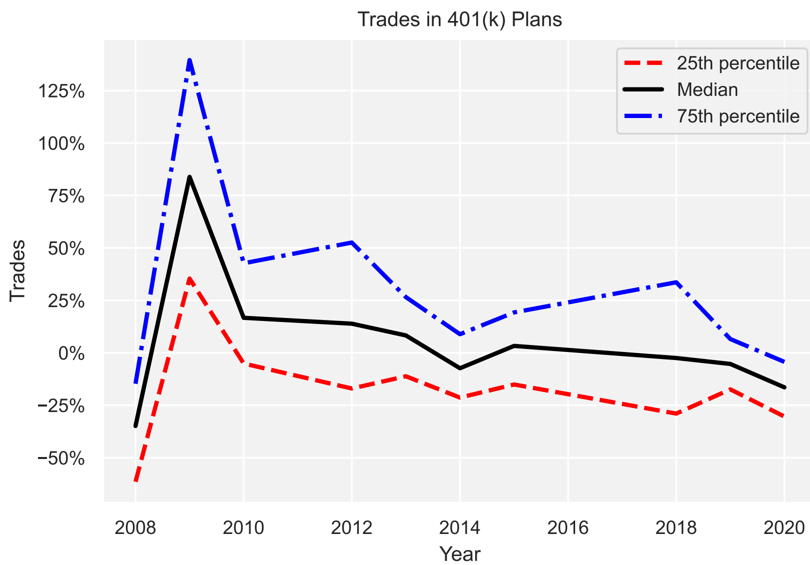


Figure 7: 401(k) Plans: Trades in Individual Stocks. This figure plots the cross-sectional distribution of 401(k) plan trading in individual stocks, defined as the percentage changes of shares in holdings by all 401(k) plans. Annual data, 2008-2020.

fixed effects.

Table 5 reports the estimates of the regression using contemporaneous returns (columns 1-3), 1-year ahead returns (columns 4-6), and cumulative 3-year ahead returns (columns 7-9). The results show that large holdings changes by 401(k) plans have a contemporaneous positive effect on individual stock returns. Quantitatively, a large position taken by 401(k) plans into a stock generates a contemporaneous annual return 12% higher than that obtained following a "normal" size trade by 401(k) plans. This evidence suggest that trading by 401(k) plans has a positive price impact on individual stocks. We also notice a return reversal of around 2% in the next year following large trading in individual stocks by 401(k) plans. This return reversal continues over the subsequent two years, but it is limited in magnitude, resulting in a permanent price impact caused by 401(k) large trades of more than 2% (e.g., 11.3%-9.5%).

Figure 7 shows the distribution of 401(k) large holdings changes across stocks, every year. Except during the Global Financial Crisis, where most 401(k) plans underweighted equities, 401(k) plans have not increased their exposure to the average stock over time. However, they do (indirectly) trade, as evidenced by the top and bottom quartile of change in holdings.

4.4 Price Impact of 401(k) Plans Demand: A Granular Instrumental Variable Approach

The previous section has documented a relation between the trading activity originated by 401(k) plans and stock returns. However, 401(k) demand for stocks is endogenous, as it is potentially related to other stock-characteristics impacting individual stock prices. To address this concern, we use the granular instrumental variable (GIV) approach of Gabaix and Koijen (2022).

Specifically, similar to Fan et al. (2022), we define the value-weighted 401(k)'s demand for individual stocks as

$$\text{Demand}_t^{401(k),VW}(n) = \sum_{i=1}^{N_t(n)} w_{i,t-1}(n) \times \frac{\text{Shares}_{i,t}(n) - \text{Shares}_{i,t-1}(n)}{\text{Shares}_{i,t-1}(n)} \tag{7}$$

where $N_t(n)$ is the total number of 401(k) plans that own stock n at time t, and the weight $w_{i,t-1}(n)$ is the fraction of stock n owned by 401(k) plan i at the end of the previous year $t-1$, e.g., the ratio of the shares of stock n held by 401(k) plan i divided by the sum of all shares of stock n cumulatively held by 401(k) plans:

$$w_{i,t}(n) = \frac{\text{Shares}_{i,t}(n)}{\sum_{j=1}^{N_t(n)} \text{Shares}_{j,t}(n)}$$

We also compute the corresponding equally-weighted demand as:

$$\text{Demand}_t^{401(k),EW}(n) = \frac{1}{N_t(n)} \sum_{i=1}^{N_t(n)} \frac{\text{Shares}_{i,t}(n) - \text{Shares}_{i,t-1}(n)}{\text{Shares}_{i,t-1}(n)} \tag{8}$$

To estimate the relationship between (value-weighted) demand originated by 401(k) plans, $\text{Demand}_t^{401(k),VW}(n)$, and individual stock returns, we run the following stock-level panel regression:

$$r_t(n) = \beta_0 + \beta_1(n) \times \left(\widehat{\text{Demand}_t^{401(k),VW}}(n)\right) + \varepsilon_t(n)$$

by instrumenting $\text{Demand}_t^{401(k),VW}(n)$ with the demand "shock" ($\text{Demand}_t^{401(k),VW}(n) -$

$\text{Demand}_t^{401(k),EW}(n)$), i.e., the difference between the value-weighted and equally-weighted flows.[27]

Table 6 reports the estimation results, controlling for size, beta, and book-to-market, stock- and time (year) fixed effects. We observe that, in the most stringent specification, the coefficient on the instrumented demand is about 0.37, suggesting that for a ten percent increase in 401(k) demand, stock prices increase by 3.7%.

5 Fund Level Channel

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

$$IO_{i,t}^{401k} = \frac{\sum_{p=1}^{M} AUM_{p,i,t}}{AUM_{i,t}} \tag{9}$$

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:

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = b_0 + \beta_{0,i} mb_t(n) + \boldsymbol{\beta_1'}\mathbf{X}_t(n) + \beta_2 IO_{i,t}^{401k} + \alpha_t + \tilde{u}_{i,t}(n) \tag{10}$$

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 Koijen 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 α_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 Koijen and Yogo (2019),

[27] As in standard IV setups, we first regress the endogeneous $\text{Demand}_t^{401(k),VW}(n)$ on the difference between the value- and equally-weighted demand (first stage), and use the (exogeneous) fitted value as regressor in the second IV stage.

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.[28] Our coefficient of interest is β_2, representing the effect of 401(k) assets on fund i demand for the average stock n.

Table 7 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).[29] In terms of magnitude, 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 column (4) and (7) of Table 7).[30] 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.

[28] We only have 268 fund-year observations, out of a total of 17,436, where the number of stocks is greater than 1,000.

[29] Table A.2 reports the same results for AUM unweighted regressions. The coefficients for mutual funds and ETFs are, respectively, 0.240 (t-stat= 4.29) and -0.060 (t-stat= -1.10), thus confirming our results.

[30] Recall that we consider only mutual funds and ETFs that are active by removing index mutual funds and ETFs.

It is also interesting to compare the fund-level results for ETFs in Table 7 to the stock-level results in Table 2. Whereas a larger fund-level 401(k) ownership does not affect the ETFs' 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 managers have flexibility in selecting their stock holdings.

The results presented in Table 7 mask some economically interesting trends. To this end, the right panel in Figure 4 shows the evolution over time of the coefficient on fund-level 401(k) ownership for both mutual funds and ETFs.[31] 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 columns (4)-(6) of Table 7). 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.

Finally, Panel B of Table 3 reports GMM estimates of the non-linear version of equation (10). 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

[31] Figure A.2 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.

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.

5.1 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:*[32]

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

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 mangers for assets with long-duration of cashflows, which the literature has found to be less risky than short duration stocks.

[32] 401(k)plan participants make periodic retirement account contributions and withdrawals, which are persistent over time. In addition, they may evaluate their present and prospective fund holdings differently due to longer investment horizons. These factors may explain the documented inertia by DC plan participants in the previous literature (see Benartzi and Thaler (2001), Madrian and Shea (2001), Choi et al. (2002), Agnew et al. (2003), Duflo and Saez (2003), Huberman and Jiang (2006), Carroll et al. (2009)) whereby retirement savers have a tendency to rebalance and trade infrequently and to follow default options.

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:*

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

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.[33] 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 $X_2, X_3, X_4, \ldots$, 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:

$$\begin{aligned} \%Share_{i,q,t+1} = \beta_1 \times \beta^{CAPM}_{q,t} + \beta_2 \times BM_{q,t} + \beta_3 \times Prof_{q,t} + \beta_4 \times Inv_{q,t} + \beta_5 \times Size_{q,t} + \\ + \beta_6 \times IO^{401k}_{i,t} + controls + u_{i,t} \end{aligned} \quad (11)$$

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 $Inv_{q,t}$

[33] 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.

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 $IO^{401k}_{i,t}$ 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 (11) 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 8 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, $IO^{401k}_{i,t}$, 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.[34] Panel B in Table 8 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

[34]This is not mechanical. The increase in portfolio weights for the bottom portfolio could have come from a reduction in the middle quintiles.

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.

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 9. 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 column (2) of Table 9 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.[35]

[35] Christoffersen and Simutin (2017) provide anecdotal evidence (from investment policy statements of

5.2 Cash Holdings of Mutual Funds and 401(k) Assets

Do active funds managing more 401(k) assets perceive their investor base to be more stable? If this were the case, one would expect funds managing a larger fraction of 401(k) assets to keep lower cash levels. In this section, we test this hypothesis by estimating the following panel regression

$$Cash_{i,t} = \alpha + \beta_1 HighIO^{401k}_{i,t} + controls_{i,t} + \varepsilon_{i,t} \tag{12}$$

where $Cash_{i,t}$ is the amount of cash as a percentage of assets of mutual fund i at time t (Chernenko and Sunderam, 2020)[36], High $IO^{401k}_{i,t}$ is a dummy equal to one if the fund level $IO^{401k}_{i,t}$ is larger than the median of $IO^{401k}_{i,t}$ in the sample, and zero otherwise. We also include standard fund controls such as lagged log fund size, expense ratio, and turnover. We focus on mutual funds, since equity ETFs might have little discretion in choosing their cash holdings.

Table 10 reports the estimation results. Columns (1)-(2) report the results using a dummy variable for 401(k) ownership, while column (3) using the continuous fund-level 401(k) ownership variable $IO^{401k}_{i,t}$.

The coefficient on the dummy is around -0.32, suggesting that mutual funds managing substantial 401(k) assets have 32% less cash holdings compared to other mutual funds. Similarly, a one percent increase in 401(k) ownership at the fund level is associated with 1.24% less mutual fund cash holdings, controlling for standard fund-level characteristics such as fund size, expense ratio, and turnover.

Overall, these results confirm our hypothesis on the stability of 401(k) flows channel, highlighting the importance of 401(k) assets in shaping investment funds' allocation decisions.

DC plans) that a large majority of DC plans list relative returns as the main criterion for investment.

[36] Cash holding of a fund is the sum of portfolio weights (as a percentage) on cash and cash equivalents in Morningstar mutual fund holdings, where cash and cash equivalents are defined by Morningstar with detail type id as 'B','BC','BD','BQ','BT','C','CA','CD','CH','CL','CQ','CR','CU','FM', 'OO','OS','P','PC', and 'Q'.

6 Conclusion

In this paper, we study the impact of 401(k) ownership on investors' demand for individual stocks. More precisely, we estimate a demand system linking 401(k) plans' ownership of both stocks and funds to the quantity and type of stocks demanded by funds. To this purpose, we introduce a new variable, stock-level 401(k) ownership, and find it to be a key determinant of investors' demand for equities.

We hypothesize that 401(k) allocations can affect stock demand in two ways. The first channel through which 401(k) plans can affect the demand of individual stocks is related to the size of stock-level 401(k) ownership. The fraction of an individual stock owned by 401(k) plans can be seen as an additional stock characteristic, similarly to book-to-market or momentum. We label this the *stock level* channel.

The second 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.

Focusing on the stock-level channel, we find that the amount of company shares owned by 401(k) plans is an important characteristic – in fact, the most important one after size – in explaining the demand of mutual funds and ETFs for a specific stock. For a one standard deviation increase in 401(k) stock ownership, the average active mutual fund demands approximately 18.6% more of the stock. The average active ETF also increases exposure to the stock by 11.5% for each standard deviation increase in 401(k) stock ownership. Most importantly, stock-level 401(k) ownership appears to be *distinct* from other forms of institutional investors, such as total mutual fund or largest (top 10) investors' ownership of a stock. After controlling for these alternative types of ownership, the magnitude of the coefficient on 401(k) ownership is barely affected, and so is its statistical significance. These results highlight the unique information content of stock-level 401(k) ownership for fund managers' decision.

We then explore the equilibrium price impact of a change in stock-level 401(k) ownership for the cross-section of stocks, over time. We estimate the institutional price pressure to be positive and increasing over our sample. We also compute the price impact for portfolios of stocks sorted on size, book-to-market, or beta (market risk), and 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. 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 (Koijen et al., 2022). To further validate the direct impact of 401(k) ownership on individual stock, we employ a matching procedure based on high and low 401(k) ownership stocks with similar characteristics, and find that stocks with positive 401(k) ownership tend to earn 3%-5% higher annual returns than similar stocks, in terms of characteristics and investor structure, not owned by 401(k) plans.

We also study the impact 401(k) plans' *trading* has on individual stock returns using two different tests. Our first methodology exploits large changes in individual stocks holdings by aggregate 401(k) plans, and find that large changes in the holdings of an individual stock substantially affect its contemporaneous return. A large positive position taken by 401(k) plans in a stock generates a 12% higher return than that implied by an average 401(k) plan trade, followed by a partial return reversal over the next two years. Our second test relies on the granular instrumental variable of Gabaix and Koijen (2022), and finds that a 10% increase in the (instrumented) demand of 401(k) plans generates an average stock price increase of 3.6%, after controlling for standard firm-specific drivers of stock returns.

As far as the fund-level channel is concerned, we document that funds managing a larger fraction of 401(k) assets display greater demand for stocks. For a standard deviation increase in 401(k) fund ownership, the average active mutual fund demands approximately 33.3% more of the average stock, while the demand from ETFs is almost unchanged. To gain further insight on the relation between 401(k) assets and fund managers' portfolio allocations, 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. Related to this point, we also find that mutual funds managing a large fraction of 401(k) assets have approximately 32% smaller cash holdings than other funds.

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

References

Agnew, Julie, Pierluigi Balduzzi, and Annika SundÃl'n (2003) "Portfolio Choice and Trading in a Large 401(k) Plan," *American Economic Review*, Vol. 93, No. 1, pp. 193–215, March.

Backus, Matthew, Christopher Conlon, and Michael Sinkinson (2021) "Common Ownership in America: 1980âĂŞ2017," *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.

Ben-David, Itzhak, Francesco Franzoni, Rabih Moussawi, and John Sedunov (2021) "The Granular Nature of Large Institutional Investors," *Management Science*, Vol. 67, No. 11, pp. 6629–6659.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song (2022a) "Discontinued Positive Feedback Trading and the Decline of Return Predictability," *Working Paper*.

——— (2022b) "Ratings-Driven Demand and Systematic Price Fluctuations," *The Review of Financial Studies*, Vol. 35, No. 6, pp. 2790–2838, 09.

Benartzi, Shlomo and Richard H. Thaler (2001) "Naive Diversification Strategies in Defined Contribution Saving Plans," *American Economic Review*, Vol. 91, No. 1, pp. 79–98, March.

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.

Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick (2009) "Optimal Defaults and Active Decisions*," *The Quarterly Journal of Economics*, Vol. 124, No. 4, pp. 1639–1674, 11.

Chen, Hsiu-Lang, Narasimhan Jegadeesh, and Russ Wermers (2000) "The Value of Active Mutual Fund Management: An Examination of the Stockholdings and Trades of Fund Managers," *The Journal of Financial and Quantitative Analysis*, Vol. 35, No. 3, pp. 343–368.

Chernenko, Sergey and Adi Sunderam (2020) "Do fire sales create externalities?" *Journal of Financial Economics*, Vol. 135, No. 3, pp. 602–628.

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2002) "Defined Contribution Pensions: Plan Rules, Participant Choices, and the Path of Least Resistance," *Tax Policy and the Economy*, Vol. 16, pp. 67–113.

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*.

Duflo, Esther and Emmanuel Saez (2003) "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, Vol. 118, No. 3, pp. 815–842, 08.

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.

Fan, Yinghua, Guanhao Feng, Scott Au, and Sayad Baronyan (2022) "Estimating Asymmetric Price Impact," *Working Paper*.

Frazzini, Andrea and Lasse Pedersen (2014) "Betting against beta," *Journal of Financial Economics*, Vol. 111, No. 1, pp. 1–25.

Gabaix, Xavier and Ralph S.J. Koijen (2022) "In Search of the Origins of Financial Fluctuations:The Inelastic Markets Hypothesis," *Working Paper*.

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*.

Huberman, Gur and Wei Jiang (2006) "Offering versus Choice in 401(k) Plans: Equity Exposure and Number of Funds," *Journal of Finance*, Vol. 61, No. 2, pp. 763–801.

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*.

Madrian, Brigitte C. and Dennis F. Shea (2001) "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *The Quarterly Journal of Economics*, Vol. 116, No. 4, pp. 1149–1187.

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, Ľuboš, 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.

Pool, Veronika K., Clemens Sialm, and Irina Stefanescu (2016) "It Pays to Set the Menu: Mutual Fund Investment Options in 401(k) Plans," *The Journal of Finance*, Vol. 71, No. 4, pp. 1779–1812.

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%
IO^{401k} (Passive Funds only)	0.85%	3.99%	9.34%	14.53%
IO^{401k} (Active Funds only)	0.28%	2.60%	7.91%	10.04%
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 collectively owned by 401(k) plans. The first row ("all funds") considers the universe of all US equity (active and index) funds. Index funds comprise mutual funds classified according to the Morningstar variables as "index funds" and "enhanced index", and ETFs with the S&P500 index as benchmark. The second and third rows, instead, only consider US equity active mutual funds and ETFs, respectively. 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. Annual data from 2007 to 2020.

	All Funds			Mutual Funds			ETFs		
Panel A: Mutual funds owned by pension plans	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$IO^{401k}_{-i,t}(n)$	0.149***	0.123***	0.110***	0.186***	0.155***	0.124***	0.115***	0.093***	0.088***
	(18.950)	(17.100)	(13.430)	(11.160)	(9.780)	(7.780)	(10.210)	(8.790)	(7.790)
Log market-to-book	0.812***	0.809***	0.814***	0.422***	0.412***	0.423***	0.964***	0.955***	0.957***
	(13.770)	(13.900)	(14.060)	(5.000)	(4.750)	(4.860)	(22.090)	(22.210)	(22.440)
Log book equity	1.411***	1.423***	1.427***	0.991***	0.996***	1.005***	1.515***	1.523***	1.523***
	(26.140)	(25.510)	(25.610)	(11.670)	(11.230)	(11.260)	(59.930)	(63.000)	(62.410)
Operating profitability	0.030***	0.0312***	0.032***	0.072***	0.073***	0.076***	0.017**	0.019***	0.020**
	(3.380)	(3.69)	(3.740)	(4.200)	(4.370)	(4.520)	(2.500)	(3.360)	(3.350)
Beta	-0.004	-0.007	-0.005	-0.020**	-0.024**	-0.021**	0.004	0.000	0.000
	(-0.930)	(-1.270)	(-1.070)	(-2.670)	(-2.730)	(-2.400)	(0.700)	(0.050)	(0.140)
Investment	0.011**	0.007	0.007	0.009	0.005	0.004	0.013**	0.008	0.008
	(2.170)	(1.340)	(1.280)	(0.760)	(0.410)	(0.350)	(2.200)	(1.520)	(1.480)
Dividend-to-book	0.000	0.003	0.002	-0.014	-0.009	-0.011	0.003	0.006	0.006
	(0.070)	(0.660)	(0.470)	(-1.390)	(-0.980)	(-1.190)	(0.390)	(0.680)	(0.650)
Top10 ownership		0.055***	0.057***		0.056***	0.058***		0.050***	0.051***
		(5.960)	(5.500)		(4.000)	(3.660)		(5.090)	(4.740)
Mutual fund ownership			0.028**			0.070***			0.012
			(2.700)			(4.310)			(1.590)

	All Funds			Mutual Funds			ETFs		
Panel B: Mutual funds not owned by pension plans	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$IO^{401k}_{-i,t}(n)$	0.106***	0.085***	0.080**	0.099***	0.079**	0.063*	0.097***	0.074***	0.081***
	(4.75)	(3.53)	(3.11)	(3.65)	(2.53)	(1.78)	(4.82)	(4.23)	(4.69)
Log market-to-book	0.761***	0.765***	0.765***	0.719***	0.719***	0.718***	0.808***	0.811***	0.810***
	(18.7)	(18.33)	(18.31)	(14.47)	(13.91)	(14.23)	(10.64)	(10.12)	(10.13)
Log book equity	1.340***	1.354***	1.354***	1.262***	1.264***	1.263***	1.443***	1.462***	1.462***
	(21.70)	(20.65)	(20.60)	(17.64)	(17.10)	(17.43)	(17.85)	(17.21)	(17.22)
Operating profitability	0.007	0.006	0.006	0.001	-0.001	-0.001	0.023*	0.026	0.026
	(0.58)	(0.43)	(0.45)	(0.11)	(-0.14)	(-0.07)	(1.81)	(1.61)	(1.59)
Beta	-0.010	-0.011	-0.011	-0.022	-0.027*	-0.026*	0.009	0.012*	0.012
	(-1.17)	(-1.18)	(-1.15)	(-1.66)	(-1.83)	(-1.78)	(1.35)	(1.78)	(1.68)
Investment	0.018***	0.017**	0.017**	0.016*	0.016	0.016	0.017***	0.017***	0.017***
	(3.10)	(2.89)	(2.86)	(1.84)	(1.74)	(1.70)	(3.77)	(3.15)	(3.19)
Dividend-to-book	-0.868	-0.743	-0.727	-2.162*	-2.032*	-1.971*	0.996*	1.020*	1.003
	(-1.17)	(-1.01)	(-0.98)	(-2.10)	(-2.02)	(-1.91)	(1.95)	(1.80)	(1.77)
Top10 ownership		0.036***	0.036***		0.027*	0.026*		0.045***	0.044***
		(3.26)	(3.24)		(1.87)	(1.82)		(3.57)	(3.47)
Mutual fund ownership			0.011			0.037*			-0.015*
			(1.17)			(2.21)			(-1.92)

Table 2: Demand system estimation - Stock level $IO^{401k}_t(n)$. This table reports estimates of the panel regression

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

where $\widehat{mb}_t(n)$ is the instrumented market-to-book equity, and $\mathbf{X}_t(n)$ includes the same variables as in Koijen and Yogo (2019), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio. $IO^{401k}_{-i,t}(n)$ 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. Columns (1)-(3) report results using all funds (including index mutual funds and ETFs), columns (4)-(6) using only active mutual funds, and columns (7)-(9) using only active ETFs. Variables are standardized to have unit-standard deviation. 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.

Panel A: Stock-level 401(k) ownership (mutual funds)			
	Coefficient	s.e.	t-stat
$IO^{401k}_{-i,t}(n)$	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
Panel B: Fund-level 401(k) ownership (mutual funds)			
	Coefficient	s.e.	t-stat
$IO^{401k}_{i,t}$	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

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

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = \exp\left\{b_0 + \beta_{0,i}\widehat{mb}_t(n) + \boldsymbol{\beta}_1'\mathbf{X}_t(n) + \beta_2 IO_t^{401k} + \alpha_t\right\}\tilde{u}_{i,t}(n)$$

where $\widehat{mb}_t(n)$ is the instrumented log market equity-to-book equity, and $\mathbf{X}_t(n)$ includes the same variables as in Koijen and Yogo (2019), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio. IO_t^{401k} indicates either the 401(k) plans ownership of the individual stock n excluding the effect through fund i (Panel A), or the 401(k) plans ownership of fund i (Panel B). We report results using only active 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.

Panel A: Characteristics of the Matched Sample		
	Treated Group	Control Group
Number of Stocks	8963	8963
Market Capitalization	11.0 B	12.2 B
Book-to-market Ratio	0.51	0.56
Beta	1.15	1.01

Panel B: Panel Regressions			
	(1)	(2)	(3)
Treated dummy$_t$	0.050***	0.042***	0.032***
	(3.352)	(3.200)	(3.248)
Size$_{t-1}$		-0.027***	-0.027***
		(-2.866)	(-3.343)
Book-to-market$_{t-1}$			-0.056
			(-1.629)
Beta$_{t-1}$			0.024
			(0.627)
Momentum$_{t-1}$			0.018
			(0.447)
Year FEs	Yes	Yes	Yes
No. Observations	17,398	17,398	17,398

Table 4: Matching Stocks: Impact of 401(k) Ownership. Panel A reports the average stock characteristics of the stocks in the treatment and control groups. Panel B reports results from of regression on the matched sample. After matching stocks as described in Section 4.2, we estimate a panel regression of annual returns of the matched pair of stocks on a treated dummy variable and various stock-level controls. Controls include lagged size, book-to-market, beta, and momentum. Standard errors are double clustered by stock and time.

	ret_t			ret_{t+1}			$ret_{t+1:t+3}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
401(k) dummy - Large Δ holdings	0.119***	0.113***	0.116***	-0.041***	-0.029***	-0.019***	-0.119***	-0.095***	-0.066***
	(4.015)	(3.849)	(4.563)	(-2.682)	(-2.650)	(-2.884)	(-4.074)	(-8.293)	(-4.270)
401(k) dummy - Small Δ holdings		-0.024	-0.033		0.049	0.039		0.098	0.092
		(-0.641)	(-0.886)		(1.124)	(0.964)		(1.128)	(1.124)
Top 10 investors dummy - Large Δ holdings			0.002			-0.018			-0.095***
			(0.100)			(-1.027)			(-2.655)
Top 10 investors dummy - Small Δ holdings			0.054***			0.068**			0.061
			(3.011)			(2.531)			(1.552)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 5: Stock-level 401(k) Trading and Returns. This table reports estimates of regressions of stock returns on changes in 401(k) plans and top 10 institutions' holdings. The dependent variables are contemporaneous returns (columns (1)-(3)), next year (t+1) returns (column (4)-(6)), and cumulative *t*+1:*t*+3 returns (column (7)-(9)). Controls include log market equity and time fixed effects. Standard errors are double clustered by stock and time.

	(1)	(2)	(3)	(4)
$\widehat{\text{Demand}_t^{401(k),VW}}(n)$	0.481***	0.401***	0.472***	0.374***
	(7.070)	(10.070)	(8.510)	(9.330)
Size_{t-1}			-0.026***	-0.291***
			(-4.390)	(-4.920)
Beta_{t-1}			-0.003	0.011
			(-0.110)	(0.570)
$\text{Book-to-market}_{t-1}$			-0.056	0.035
			(-1.320)	(0.730)
Momentum_{t-1}			-0.027	-0.070*
			(-0.650)	(-1.950)
Stock FEs	No	Yes	No	Yes
Year FEs	Yes	Yes	Yes	Yes

Table 6: Granular Instrumental Variable Regression. This table reports estimates of the GIV stock-level panel regression

$$r_t(n) = \beta_0 + \beta_1(n) \times \left(\widehat{\text{Demand}_t^{401(k),VW}}(n)\right) + \varepsilon_t(n)$$

The dependent variable is annual stock returns in year t. The variable of interested is 401(k) plans' demand, instrumented by GIV $\widehat{\text{Demand}_t^{401(k),VW}}(n)$. Standard errors are double clustered by stock and time.

	All Funds			Mutual Funds			ETFs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$IO^{401k}_{i,t}$	0.169**	0.160**	0.157**	0.333**	0.328**	0.319**	0.011	0.008	0.008
	(2.390)	(2.260)	(2.240)	(2.910)	(2.890)	(2.830)	(0.110)	(0.080)	(0.090)
Log market-to-book	0.766***	0.750***	0.758***	0.637***	0.608***	0.610***	0.560***	0.540***	0.552***
	(7.340)	(7.250)	(7.390)	(9.510)	(9.130)	(9.120)	(4.540)	(4.430)	(4.590)
Log book equity	1.289***	1.331***	1.348***	1.032***	1.088***	1.112***	1.105***	1.117***	1.132***
	(14.170)	(14.170)	(14.540)	(11.050)	(11.350)	(11.740)	(9.930)	(9.790)	(10.000)
Operating profitability	0.036**	0.043***	0.048	0.031*	0.043**	0.054***	0.042**	0.049**	0.052**
	(2.800)	(3.270)	(3.630)	(1.810)	(2.510)	(3.250)	(2.660)	(2.890)	(2.910)
Beta	-0.033**	-0.043**	-0.035**	-0.042**	-0.050***	-0.041**	-0.064***	-0.079***	-0.073***
	(-2.660)	(-3.050)	(-2.580)	(-2.910)	(-3.340)	(-2.870)	(3.520)	(4.360)	(-4.060)
Investment	-0.013	-0.021	-0.022	-0.021	-0.029	-0.030*	-0.031*	-0.038**	-0.039**
	(-0.940)	(-1.390)	(-1.510)	(-1.280)	(-1.720)	(-1.830)	(2.030)	(2.530)	(-2.570)
Dividend-to-book	-0.047***	-0.030**	-0.023*	-0.090***	-0.065***	-0.054***	0.035	0.040*	0.045*
	(-3.600)	(-2.510)	(-1.870)	(-5.970)	(-4.770)	(-4.000)	(1.530)	(1.900)	(2.160)
Top10 ownership		0.136***	0.121***		0.162***	0.137***		0.081**	0.071**
		(7.060)	(6.280)		(7.680)	(6.370)		(3.160)	(2.810)
Mutual Fund ownership			0.142***			0.183***			0.115***
			(7.420)			(7.870)			(4.850)

Table 7: Demand system estimation - Fund level $IO^{401k}_{i,t}$. This table reports estimates of the panel regression

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

where $\widehat{mb}_t(n)$ is the instrumented log market equity-to-book equity, and $\mathbf{X}_t(n)$ includes the same variables as in Koijen and Yogo (2019), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio. $IO^{401k}_{i,t}$ is the 401K plans ownership of fund i, and α_t are time fixed effects. The funds in the regressions are AUM-weighted. Columns (1)-(3) report results using all funds (including index mutual funds and ETFs), columns (4)-(6) using only active mutual funds, and columns (7)-(9) using only active ETFs. Variables are standardized to have unit-standard deviation. 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.

Panel A: Beta			
	High (Low Beta)	Low (High Beta)	High-Low
	(1)	(2)	(3)
$IO^{401k}_{i,t}$	-0.027***	0.015***	-0.031***
	(0.006)	(0.004)	(0.009)
weight$_t$	0.883***	0.872***	0.897***
	(0.020)	(0.016)	(0.015)

Panel B: Duration			
	High (Short Duration)	Low (Long Duration)	High-Low
	(1)	(2)	(3)
$IO^{401k}_{i,t}$	-0.028***	0.008**	-0.023***
	(0.007)	(0.003)	(0.006)
weight$_t$	0.890***	0.863***	0.900***
	(0.009)	(0.011)	(0.009)

Panel C: Momentum			
	High (Winner)	Low (Losers)	High-Low
	(1)	(2)	(3)
$IO^{401k}_{i,t}$	0.021**	-0.006	0.028**
	(0.010)	(0.006)	(0.014)
weight$_t$	0.869***	0.808***	0.859***
	(0.039)	(0.019)	(0.034)

Panel D: Size			
	High (Small)	Low (Large)	High-Low
	(1)	(2)	(3)
$IO^{401k}_{i,t}$	0.002	-0.015**	0.014
	(0.005)	(0.007)	(0.008)
weight$_t$	0.930***	0.996***	1.000***
	(0.023)	(0.003)	(0.005)

Table 8: 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 $t+1$, $w^P_{i,t+1}$ for $P = High, Low, High - Low$, on the 401(k) plans ownership of fund i, $IO^{401k}_{i,t}$, controlling for lagged portfolio weights, $w^P_{i,t}$, 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 α_{t+1}	
	(1)	(2)	(3)	(4)
Fund β_t	1.918***	1.851**	-0.458***	-0.465***
	(2.755)	(2.497)	(-6.472)	(-5.976)
Log fund size$_t$	0.065	0.040	0.029***	0.030***
	(0.995)	(0.617)	(4.089)	(3.623)
Expenses$_t$	0.742**	0.798**	0.072**	0.074*
	(2.196)	(2.350)	(2.171)	(1.681)
Relative return$_t$	-0.014	-0.015	0.000	0.000
	(-0.621)	(-0.657)	(0.263)	(0.238)
Turnover$_t$	-0.223	-0.223	0.003	0.003
	(-1.595)	(-1.594)	(0.110)	(0.109)
Amihud illiquidity$_t$	0.052	0.057	0.001	0.001
	(1.156)	(1.218)	(0.231)	(0.283)
$IO^{401k}_{i,t}$		1.938***		0.095
		(3.000)		(0.984)

Table 9: 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 β 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. $IO^{401k}_{i,t}$ 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.

	(1)	(2)	(3)
High $IO^{401k}_{i,t}$ (dummy)	-0.322**	-0.322**	
	(-2.048)	(-2.495)	
$IO^{401k}_{i,t}$			-1.240**
			(-2.117)
Log fund size$_t$	-0.001	-0.012	-0.020
	(-0.039)	(-0.461)	(-0.450)
Expense$_t$	2.270***	2.303***	-0.263
	(7.252)	(5.143)	(-0.268)
Turnover$_t$	-0.367*	-0.376*	0.024
	(-1.804)	(-1.838)	(0.492)
Fund FEs	No	No	Yes
Year FEs	No	Yes	Yes

Table 10: Mutual Funds Cash Holdings.This table reports the estimates of the following fund-year panel regression:

$$Cash_{i,t} = \alpha + \beta_1 HighIO^{401k}_{i,t} + controls_{i,t} + \varepsilon_{i,t}$$

The dependent variable is mutual fund cash holdings at the end of year t, which is the sum of portfolio weights (as a percentage) on cash and cash equivalents in Morningstar mutual fund holdings. High $IO^{401k}_{i,t}$ is a dummy equal to one if the fund level $IO^{401k}_{i,t}$ is larger than the median of $IO^{401k}_{i,t}$ in the sample, and zero otherwise. Log fund size is the logarithm of the fund AUM. Expenses is net expense ratio (as a percentage), 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. Column (3) reports the result using the stock-level 401(k) ownership variable $IO^{401k}_{i,t}$. Standard errors are double clustered by fund and year.

Internet Appendix A

A.1 Coefficients on Other Characteristics

Figure A.2 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 (10)). 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.1 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 (4)). Comparing Figure A.2 to Figure A.1, 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 (10) and (4)).

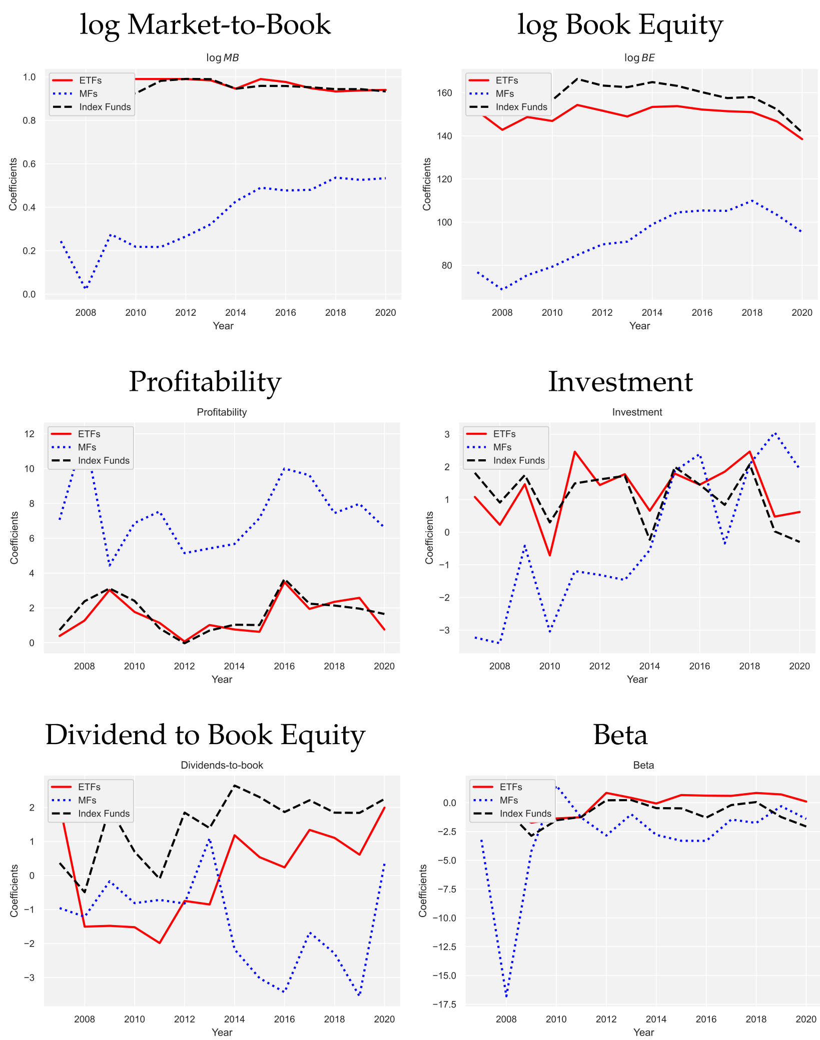


Figure A.1: Coefficients on the other characteristics - Stock level. This figure shows the annual coefficients in (4), separately for mutual funds, ETFs, and index funds, estimated by pooled OLS using assets under management as weights. The regression is estimated year by year. Except for log market-to-book 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.

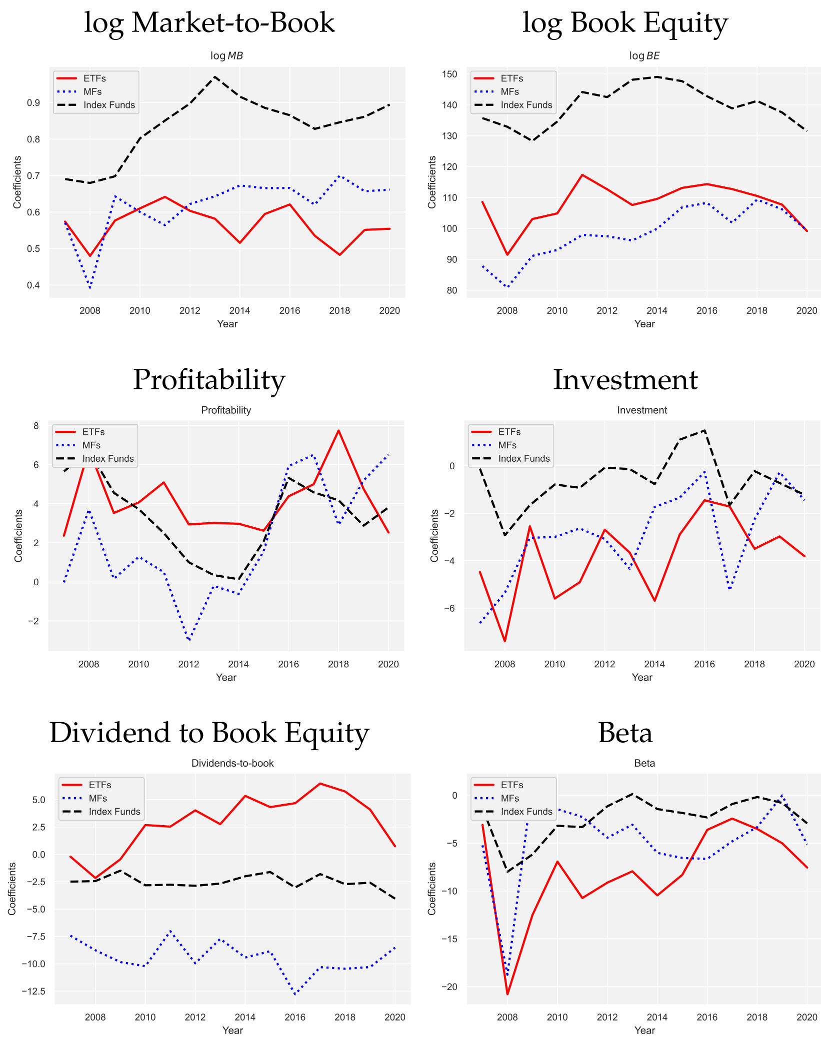


Figure A.2: Coefficients on the other characteristics - Fund level. This figure shows the annual coefficients in (10), separately for mutual funds, ETFs, and index funds, estimated by pooled OLS using assets under management as weights. The regression is estimated year by year. Except for log market-to-book 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 endogeneous 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 Koijen 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 7, controlling for these additional fund-specific variables: fund size, fund style, and fund family fixed effects.[37] 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 exogeneous. 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 Koijen and Yogo (2019). We save the residuals $\widehat{\varepsilon}_{i,t}$ from this estimation. Next, we regress our fund-level 401(k) variable $IO_{i,t}^{401k}$ on

[37] The *fund style* and *fund family* variables are from Morningstar.

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

By construction, $\widehat{\eta}_{i,t}$ is the component of fund-level 401(k) ownership $IO_{i,t}^{401k}$ that is orthogonal to $\widehat{\varepsilon}_{i,t}$, e.g., the residual from the original model. In other words, it is exogeneous by construction. Lastly, we re-estimate (1) using $\widehat{\eta}_{i,t}$ (instead of $IO_{i,t}^{401k}$) as regressor. The coefficient on $\widehat{\eta}_{i,t}$ 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 exogeneous.

Panel A: Mutual funds owned by pension plans									
	All Funds			**Mutual Funds**			**ETFs**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$IO^{401k}_{-i,t}(n)$	0.116***	0.106***	0.099***	0.122***	0.117***	0.106***	0.082***	0.068***	0.066***
	(17.06)	(16.47)	(14.59)	(11.07)	(10.96)	(9.64)	(7.80)	(7.56)	(6.82)
Log market-to-book	0.580***	0.574***	0.576***	0.264***	0.254***	0.258***	0.784***	0.785***	0.786***
	(12.25)	(11.68)	(11.73)	(4.58)	(4.27)	(4.31)	(13.39)	(13.37)	(13.44)
Log book equity	1.117***	1.119***	1.121***	0.751***	0.745***	0.748***	1.326***	1.340***	1.341***
	(27.14)	(25.65)	(25.71)	(14.52)	(13.85)	(13.89)	(28.76)	(28.38)	(28.41)
Operating profitability	0.046***	0.048***	0.049***	0.066***	0.068***	0.069***	0.034***	0.036***	0.036***
	(8.03)	(7.87)	(7.91)	(8.6)	(7.9)	(7.97)	(3.73)	(3.85)	(3.85)
Beta	-0.014**	-0.018**	-0.017**	-0.031***	0.036***	-0.035**	0.006	0.002	0.002
	(-2.8)	(-2.72)	(-2.64)	(-3.56)	(-3.2)	(-3.13)	(1.41)	(0.51)	(0.58)
Investment	0.001	-0.002	-0.002	-0.006	0.010*	-0.010*	0.015**	0.012*	0.012*
	(0.33)	(-0.48)	(-0.52)	(-1.07)	(-1.85)	(-1.9)	(2.54)	(1.87)	(1.86)
Dividend-to-book	0.008	0.007	0.007	0.004	0.003	0.003	0.013	0.015	0.014
	(1.64)	(1.7)	(1.61)	(0.72)	(0.64)	(0.52)	(1.33)	(1.52)	(1.51)
Top10 ownership		0.023***	0.024***		0.003	0.004		0.037***	0.037***
		(3.85)	(3.79)		(0.52)	(0.53)		(4.17)	(4.14)
Mutual Fund ownership			0.016**			0.026***			0.006
			(2.8)			(3.66)			(1.01)

Panel B: Mutual funds not owned by pension plans									
	All Funds			**Mutual Funds**			**ETFs**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$IO^{401k}_{-i,t}(n)$	0.106***	0.085***	0.080**	0.099***	0.079**	0.063*	0.097***	0.074***	0.081***
	(4.75)	(3.53)	(3.11)	(3.65)	(2.53)	(1.78)	(4.82)	(4.23)	(4.69)
Log market-to-book	0.761***	0.765***	0.765***	0.719***	0.719***	0.718***	0.808***	0.811***	0.810***
	(18.7)	(18.33)	(18.31)	(14.47)	(13.91)	(14.23)	(10.64)	(10.12)	(10.13)
Log book equity	1.340***	1.354***	1.354***	1.262***	1.264***	1.263***	1.443***	1.462***	1.462***
	(21.70)	(20.65)	(20.60)	(17.64)	(17.10)	(17.43)	(17.85)	(17.21)	(17.22)
Operating profitability	0.007	0.006	0.006	0.001	-0.001	-0.001	0.023*	0.026	0.026
	(0.58)	(0.43)	(0.45)	(0.11)	(-0.14)	(-0.07)	(1.81)	(1.61)	(1.59)
Beta	-0.010	-0.011	-0.011	-0.022	-0.027*	-0.026*	0.009	0.012*	0.012
	(-1.17)	(-1.18)	(-1.15)	(-1.66)	(-1.83)	(-1.78)	(1.35)	(1.78)	(1.68)
Investment	0.018***	0.017**	0.017**	0.016*	0.016	0.016	0.017***	0.017***	0.017***
	(3.10)	(2.89)	(2.86)	(1.84)	(1.74)	(1.70)	(3.77)	(3.15)	(3.19)
Dividend-to-book	-0.868	-0.743	-0.727	-2.162*	-2.032*	-1.971*	0.996*	1.020*	1.003
	(-1.17)	(-1.01)	(-0.98)	(-2.10)	(-2.02)	(-1.91)	(1.95)	(1.80)	(1.77)
Top10 ownership		0.036***	0.036***		0.027*	0.026*		0.045***	0.044***
		(3.26)	(3.24)		(1.87)	(1.82)		(3.57)	(3.47)
Mutual fund ownership			0.011			0.037*			-0.015*
			(1.17)			(2.21)			(-1.92)

Table A.1: Demand system estimation - Stock level $IO^{401k}_t(n)$, observations not AUM-weighted. This table reports estimates of the panel regression

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

where $\widehat{mb}_t(n)$ is the instrumented market-to-book equity, and $\mathbf{X}_t(n)$ includes the same variables as in Koijen and Yogo (2019), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio. $IO^{401k}_{-i,t}(n)$ 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 not AUM-weighted. Columns (1)-(3) report results using all funds (including index mutual funds and ETFs), columns (4)-(6) using only active mutual funds, and columns (7)-(9) using only active ETFs. Variables are standardized to have unit-standard deviation. 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.

	All Funds			Mutual Funds			ETFs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$IO^{401k}_{i,t}$	0.168***	0.167***	0.165***	0.240***	0.232***	0.228***	-0.060	-0.054	-0.053
	(3.38)	(3.31)	(3.28)	(4.29)	(4.1)	(4.06)	(-1.1)	(-1.01)	(-1.02)
Log market-to-book	0.457***	0.444***	0.452***	0.513***	0.490***	0.492***	0.416***	0.422***	0.431***
	(7.83)	(7.66)	(7.84)	(9.08)	(8.9)	(8.98)	(4.36)	(4.53)	(4.67)
Log book equity	0.949***	0.974***	0.995***	0.810***	0.834***	0.851***	0.961***	0.985***	0.999***
	(19.56)	(18.86)	(19.27)	(13.27)	(12.81)	(12.93)	(12.01)	(11.98)	(12.2)
Operating profitability	0.074***	0.083***	0.089***	0.033***	0.042***	0.049***	0.067***	0.069***	0.072***
	(8.87)	(9.21)	(9.73)	(3.6)	(4.41)	(5.32)	(4.13)	(4.00)	(4.15)
Beta	-0.064***	-0.075***	-0.067***	-0.053***	-0.061***	-0.054***	-0.070***	-0.079***	-0.074***
	(-5.69)	(-5.4)	(-4.88)	(-4.19)	(-4.05)	(-3.73)	(-3.84)	(-4.14)	(-3.87)
Investment	-0.067***	-0.076***	-0.077***	-0.061***	-0.070***	-0.070***	-0.038**	-0.044***	-0.044***
	(-6.6)	(-8.15)	(-8.34)	(-6.38)	(-8.6)	(-8.78)	(-3.02)	(-3.41)	(-3.48)
Dividend-to-book	-0.059***	-0.052***	-0.046***	-0.090***	-0.081***	-0.075***	0.007	0.011	0.013
	(-5.3)	(-4.94)	(-4.46)	(-8.56)	(-8.49)	(-7.87)	(0.34)	(0.53)	(0.64)
Top10 ownership		0.073***	0.056***		0.078***	0.059***		0.051**	0.043**
		(5.26)	(3.71)		(6.05)	(4.08)		(2.81)	(2.38)
Mutual Fund ownership			0.141***			0.124***			0.082***
			(11.68)			(10.54)			(5.71)

Table A.2: Demand system estimation - Fund level $IO^{401k}_{i,t}$, observations not AUM-weighted. This table reports estimates of the panel regression

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

where $\widehat{mb}_t(n)$ is the instrumented log market equity-to-book equity, and $\mathbf{X}_t(n)$ includes the same variables as in Koijen and Yogo (2019), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio. $IO^{401k}_{i,t}$ is the 401K plans ownership of fund i, and α_t are time fixed effects. The funds in the regressions are not AUM-weighted. Columns (1)-(3) report results using all funds (including index mutual funds and ETFs), columns (4)-(6) using only active mutual funds, and columns (7)-(9) using only active ETFs. Variables are standardized to have unit-standard deviation. 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)
$IO^{401k}_{i,t}$	0.23***	0.26***	0.18**
	(3.42)	(3.29)	(2.33)
Log market-to-book	0.75***	0.70***	0.64***
	(7.89)	(6.87)	(6.25)
Log book equity	1.27***	1.21***	1.18***
	(15.29)	(13.32)	(13.59)
Operating Profitability	0.03**	0.04**	0.05***
	(2.66)	(2.84)	(4.09)
Beta	-0.03**	-0.03**	-0.03**
	(-2.87)	(-2.60)	(-2.83)
Investment	-0.01	-0.01	-0.01
	(-1.07)	(-1.20)	(-0.74)
Dividend-to-book	-0.04**	-0.04***	-0.02
	(-2.96)	(-3.01)	(-1.59)
Fund size	-0.40***	-0.30***	-0.22***
	(-5.87)	(-4.24)	(-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 $IO^{401k}_{i,t}$. This table reports estimates of the panel regression

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

where $\widehat{mb}_t(n)$ is the instrumented log market equity-to-book equity, and $\mathbf{X}_t(n)$ includes the same variables as in Koijen and Yogo (2019), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio. $IO^{401k}_{i,t}$ is the 401K plans ownership of fund i, and α_t are time (year) fixed effects. We include additional fund-level controls to rule out potential endogeneity concerns. $Fundsize_{i,t}$ 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 relies on data from Morningstar, which provides detailed holdings of *individual* mutual funds and ETFs. Instead, the analysis in Section 4.1 relies on the Thomson Reuters' s34 file, which provides aggregated holdings of *all funds* under the managerâĂŹs control (Koijen and Yogo (2019), Koijen et al. (2022)).

Table B.1 repeats the same analysis presented in Table 2 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 column (4) of Table 2, 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

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

where $\widehat{mb}_t(n)$ is the instrumented market-to-book equity, and $\mathbf{X}_t(n)$ includes the same variables as in Koijen 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.

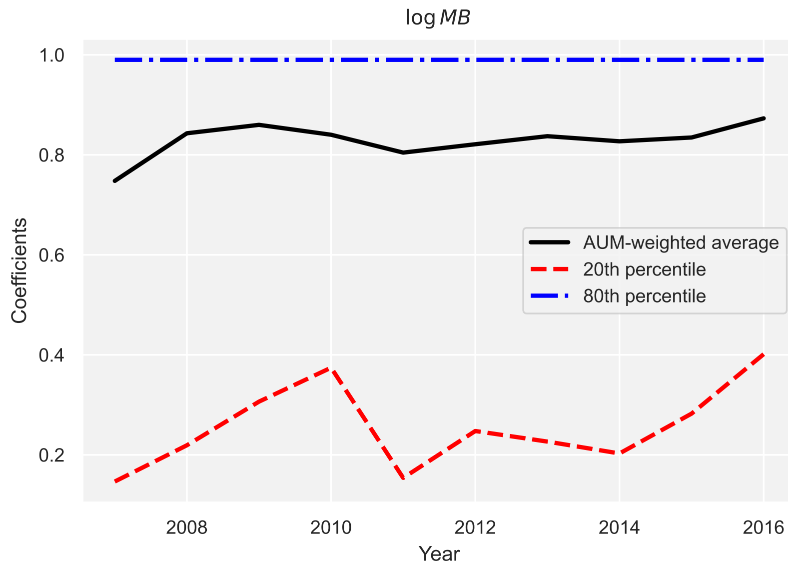
The coefficients on the other characteristics, e.g., beta, investment, and dividend-to-book are small in both datasets. Overall, it appears that the empirical results using the s34 dataset are in line with those reported in Section 4 and, thus, the analysis in Section 4.1 is informative of the equilibrium price impact of a change in 401(k) stock-level ownerhship.

B.2 Holdings scraped directly from 13F filings

In this section, we repeat our computation of the equilibrium price impact presented in Section 4.1 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 Koijen 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.

Panel A: Log Market-to-Book



Panel B: Stock-level 401(k) Ownership

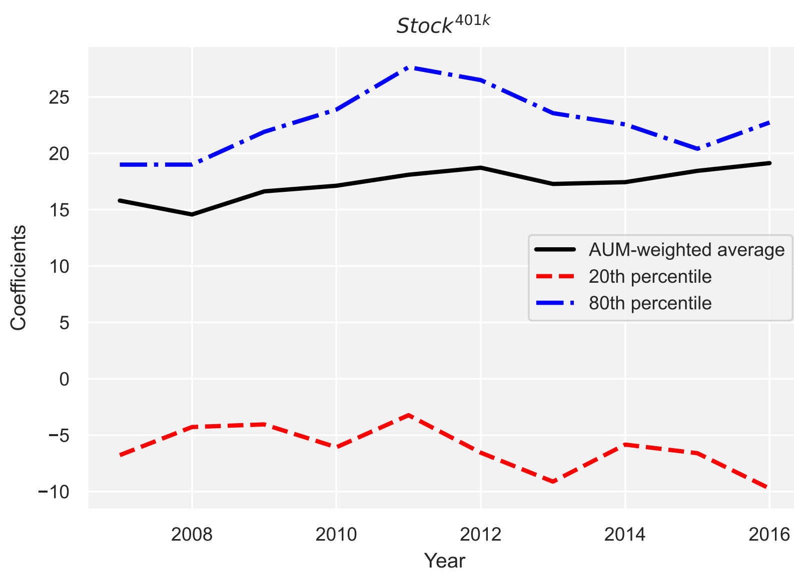


Figure B.1: Price impact: relevant coefficients. This figure shows the annual coefficient on log market-to-book (top panel) and stock-level 401(k) ownership (bottom panel) for financial institutions in Backus, Conlon and Sinkinson (2021), estimated annually, by GMM with zero weights. Variables are standardized (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 (6) 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 ≥ 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 Koijen 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(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 Koijen et al. (2022), we impose the economic constraint log(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 Koijen 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 Koijen 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).

	All Funds			Mutual Funds			ETFs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$IO^{401k}_{-i,t}(n)$	0.133***	0.107***	0.096***	0.179***	0.140***	0.107***	0.116***	0.093***	0.088***
	(19.89)	(25.71)	(16.79)	(10.00)	(7.62)	(5.24)	(9.63)	(8.32)	(7.36)
Log market-to-book	0.939***	0.934***	0.939***	0.441***	0.429***	0.440***	0.961***	0.953***	0.955***
	(18.47)	(18.40)	(18.62)	(5.16)	(4.78)	(4.90)	(20.88)	(21.28)	(21.49)
Log book equity	1.572***	1.580***	1.583***	0.983***	0.994***	1.002***	1.518***	1.524***	1.525***
	(25.06)	(25.08)	(25.12)	(11.86)	(11.32)	(11.40)	(63.65)	(66.93)	(66.38)
Operating profitability	0.017**	0.018**	0.019**	0.065***	0.065***	0.068***	0.019**	0.022***	0.023***
	(2.92)	(2.90)	(2.97)	(4.01)	(4.29)	(4.46)	(2.73)	(3.57)	(3.55)
Beta	-0.003	-0.006	-0.004	-0.022**	-0.026**	-0.023*	0.002	-0.002	-0.001
	(-0.91)	(-1.52)	(-1.25)	(-2.41)	(-2.53)	(-2.23)	(0.30)	(-0.25)	(-0.15)
Investment	0.010**	0.009*	0.008*	0.000	-0.005	-0.005	0.010	0.006	0.006
	(2.95)	(1.86)	(1.81)	(-0.01)	(-0.33)	(-0.37)	(1.63)	(1.02)	(0.98)
Dividend-to-book	0.003	0.005	0.004	-0.024	-0.018	-0.019	0.006	0.007	0.007
	(0.60)	(1.20)	(0.99)	(-1.7)	(-1.27)	(-1.36)	(0.69)	(0.79)	(0.76)
Top10 ownership		0.054***	0.056***		0.075***	0.077***		0.049***	0.050***
		(6.54)	(5.93)		(3.68)	(3.46)		(5.21)	(4.79)
Mutual fund ownership			0.023**			0.074***			0.013
			(2.63)			(3.73)			(1.55)

Table C.1: Demand system estimation - Stock level $IO^{401k}_{t}(n)$ TNA unwinsorized. This table reports estimates of the panel regression

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = b_0 + \beta_{0,i}\widehat{mb}_t(n) + \boldsymbol{\beta}_1'\mathbf{X}_t(n) + \beta_2 IO^{401k}_{-i,t}(n) + \alpha_{i,t} + \tilde{u}_{i,t}(n)$$

where $\widehat{mb}_t(n)$ is the instrumented market-to-book equity, and $\mathbf{X}_t(n)$ includes the same variables as in Koijen and Yogo (2019), i.e., log book equity, operating profitability, stock market beta, investment, and dividend-to-book ratio. $IO^{401k}_{-i,t}(n)$ 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. Columns (1)-(3) report results using all funds (including index mutual funds and ETFs), columns (4)-(6) using only active mutual funds, and columns (7)-(9) using only active ETFs. Variables are standardized to have unit-standard deviation. 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.