NBER WORKING PAPER SERIES

WHAT DRIVES VARIATION IN INVESTOR PORTFOLIOS?
EVIDENCE FROM RETIREMENT PLANS

Mark L. Egan
Alexander MacKay
Hanbin Yang

Working Paper 29604
http://www.nber.org/papers/w29604

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021, Revised June 2022

We thank John Beshears, John Campbell, Xavier Gabaix, Sam Hanson, David Laibson, Erik Loualiche, Andrei Shleifer, Adi Sunderam, Motohiro Yogo, and the seminar participants at Harvard Business School, Indiana University, Johns Hopkins Carey Business School, NBER Asset Pricing Meetings, Oklahoma State University, Southern Methodist University, the University of Minnesota Corporate Finance Conference, the University of Texas, and the University of Toulouse. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

What Drives Variation in Investor Portfolios? Evidence from Retirement Plans
Mark L. Egan, Alexander MacKay, and Hanbin Yang
NBER Working Paper No. 29604
December 2021, Revised June 2022
JEL No. G0,G11,G12,G40,G5,G51,J32

ABSTRACT

We document new patterns in investment behavior using a comprehensive dataset of 401(k) plans from 2009 through 2019. We show that there is substantial heterogeneity in asset allocation across plans, and that these differences are systematically predictable by sector of employment and demographic characteristics. For example, higher income and education is associated with more exposure to equities, while a greater share of minorities and retirees is associated with less equity exposure. These patterns cannot be rationalized by differences in investment options or plan participation. To understand observed investment behavior, we use a revealed preference approach that allows us to recover heterogeneity in investors' (subjective) expectations and risk preferences. We find that differences in expectations play an important role in explaining portfolios. Further, we show that investors appear to form expectations based on local sources of information such as county-level GDP growth, home prices, and employer past performance. Overall, our findings are consistent with a model in which heterogeneity in investor expectations reflects idiosyncratic experiences and local environments.

Mark L. Egan
Harvard Business School
Baker Library 365
Soldiers Field
Boston, MA 02163
and NBER
megan@hbs.edu

Alexander MacKay
Harvard Business School
Soldiers Field
Morgan 217
Boston, MA 02163
amackay@hbs.edu

Hanbin Yang
Harvard University
hanbin.v.yang@gmail.com

1 Introduction

To what extent do individuals make different investment allocation decisions, and what drives those differences? Understanding these dimensions of investor behavior is important for the regulation of financial markets, the measurement of macroeconomic impacts, and informing policy decisions such as the design of retirement plans.

In this paper, we address these questions using a comprehensive dataset of portfolio allocations within employer-sponsored retirement plans. This novel dataset represents the near universe of 401(k) plans from 2009 through 2019. Over this period, plan participation rates are high—74 percent on average—and there are substantial differences in investment behavior across plans. On average, 44 percent of assets are allocated to US equity funds, but this ranges from 17 percent to 64 percent for the 10th and 90th percentiles. We examine the factors driving these portfolio allocations and show how we can use observed portfolio allocations to learn about investors' subjective beliefs and preferences.

We make three distinct contributions. First, we document reduced-form differences in portfolio allocations and how these differences correlate with investor characteristics, such as demographics and employment. Second, we develop an empirical model of portfolio allocation that allows us to nonparametrically recover idiosyncratic beliefs and risk preferences. Third, we use the estimates from the model to shed new light on the cross-sectional and time-series drivers of investor beliefs. Overall, our results suggest that local information—through demographic, geographic, and employment channels—can drive substantial differences in beliefs and investor behavior.

We study portfolio allocations in the context of defined contribution plans, which are employer-sponsored (and tax-advantaged) investment accounts. Approximately half of Americans participate in the stock market, and for 60 percent of those participants, defined contribution plans are their sole source of equity exposure (Badarinza et al., 2016).[1] We focus on 401(k) plans, which are the most prevelant type of defined contribution plan. As of 2021, Americans held roughly $7 trillion in 401(k) assets.[2] Our data on 401(k) plan allocations comes from BrightScope Beacon.[3] BrightScope Beacon provides annual plan-level details about investment menus and fund allocations for 70,000 different 401(k) plans over the period 2009-2019. The entire BrightScope dataset covers 85 percent of assets in ERISA defined contribution plans. We merge our 401(k) data with return and fee from CRSP, and industry-by-county-by-year level demographic proxies from American Community Survey (ACS).

Using this data, we document substantial heterogeneity in allocations across plans, and

[1]Defined contribution plans account for the bulk of equity participation in the US and roughly one third of retirement assets. https://www.ici.org/system/files/2021-06/21_rpt_recsurveyq1.pdf

[2]https://www.ici.org/faqs/faq/401k/faqs_401k

[3]Previous work has used a subset of this data and has primarily focused on plan design and its impact on 401(k) participation. In contrast, we observe plan level data for 85% of 401(k) assets over 10 years to study investment allocations across plans.

we find that plan allocations are highly correlated with participant demographics. Plans with wealthier and more highly educated participants tend to have higher equity exposure, while plans with a greater share of older, retired, and minority participants tend to have lower equity exposures. These differences cannot be explained by differences in retirement-plan participation among these groups, as we study the within-plan allocation decisions conditional on participation. Nor can they be explained by the composition of the menu, as plan menus are largely uncorrelated with participant demographics.[4] Instead, our analysis suggests that investors make conscious (and different) allocation decisions. We estimate that a 10 basis point (bp) increase in fund expense ratios is associated with a 6.7% decrease in demand, which suggests that fees play an important role in allocation decisions (consistent with the evidence documented in Kronlund et al., 2021). We also find variation in 401(k) holdings over time: adjusting for returns, the one-year autocorrelation in fund holdings is 0.77-0.89, which indicates that some investors actively rebalance their portfolios.

Overall, we find that there is substantial variation in portfolio holdings across plans, that this variation is correlated with participant demographics, and that little of the variation is explained by differences in plan menus. Consequently, we focus on how differences in risk preferences and beliefs across investors explain variation in holdings. To interpret the decisions of investors, we model an investor's portfolio decision as a mean-variance optimization problem. When forming a portfolios, an investor trades off their subjective expectations with the corresponding additional risk, which is scaled by their risk preference.

We implement a new identification strategy to nonparametrically recover the joint distributions of beliefs and risk preferences across investors. Within our portfolio allocation model, these primitives are identified by exploiting exogenous variation in fund fees. By understanding how investors would re-allocate in response to a change in fees, which shift expected returns, we can measure how investors trade-off risk and returns and consequently recover investors' subjective beliefs about expected returns. We use Hausman-type instruments to ensure that the variation in expenses we exploit is orthogonal to investor beliefs (Hausman, 1996). We interpret the recovered risk preference as risk aversion, and we discuss the limitations of our approach.

We use the model to estimate the time-varying distributions of risk aversion and expected returns for each investment option, which may vary arbitrarily across plans. Our estimates reflect the average preferences and beliefs of investors within each employer-sponsored plan, since we only observe average portfolio allocations in our data. For example, in 2019, the average participant in the IBM 401(k) Plus Plan held 4% of their portfolio in the Vanguard Russell 1000 Value Index and 2% in the Vanguard European Stock Index. Using our framework,

[4] While some earlier work indicates that the choices of investors are driven by the menu of funds (Benartzi and Thaler, 2001, 2007), we find a substantially weaker relationship that is more consistent with the evidence in Huberman and Jiang (2006). One potential reason is that 401(k) sponsors have substantially increased the number of options in fund menus over the past 30 years, which has made the menus themselves less important.

we can then separately recover the average IBM 401(k) participant's expectations about the return of both the Vanguard Russell 1000 Value Index and the Vanguard European Stock Index as of 2019.

We recover reasonable time-varying distributions of both risk aversion and beliefs that are consistent with previous research and realized returns. In our baseline specification, we estimate an average constant relative risk aversion parameter of roughly 4, which is comparable to what other researchers have found in the literature.[5] The average investor in our sample behaved as if she expected the excess return of the market to be 9.6% over the period 2009-2019. To put this in perspective, the compound annual excess return of the of the S&P 500 was 10.7% over the same period. We also find very similar estimates of risk aversion and expected returns if we estimate the model using data from newly introduced plans in the year of inception and exclude default options, such that the allocation decisions are more likely to reflect the active choices of plan participants.

We find that accounting for heterogeneity in both risk aversion and beliefs is important for fitting the investment patterns we find in the data. A simple two-parameter model with risk aversion and beliefs explains more than 50% of the reduced-form variation in equity holdings across plans. To more precisely evaluate the extent to which heterogeneity in beliefs and risk aversion shape investment behavior, we use our model to calculate counterfactual allocations where investors have identical beliefs, identical risk aversion, or both. We find that heterogeneity in beliefs contributes to the majority of variation in across-plan allocations.

With the estimates in hand, we explore how beliefs and risk preferences depend systematically on observable characteristics. We find that wealthier and more educated investors tend to have more optimistic market expectations, which is consistent with previous experimental and survey evidence documenting that households with lower socioeconomic status are more pessimistic about future stock returns and macroeconomic conditions (Kuhnen and Miu, 2017; Das et al., 2020). Conversely, older and minority investors tend to have more pessimistic market expectations. We also find that investors' beliefs are correlated with their work experience. For example, investors working in the real estate sector are 27% (2.3 pp) more optimistic about the expected return of the market than investors working in the construction sector, despite both sectors having potentially similar risk exposures.

Risk aversion also varies with demographics and employment. Older and more educated investors behave as if they are more risk averse while wealthier investors, as measured by income, appear more risk tolerant. The variation in risk aversion and beliefs provides insight into why equity exposure varies with investor demographics. For example, our results suggest that beliefs, rather than risk aversion, explain why educated investors tend to tilt their portfolios towards equities. Conversely, both risk aversion and beliefs help explain why older investors

[5]For example, using life cycle models, Fagereng et al. (2017) estimate relative risk aversion of 7.3, Calvet et al. (2019) estimate relative risk aversion of 5.8, Meeuwis (2019) estimate relative risk aversion of 5.4, and Choukhmane and de Sliva (2022) estimate relative risk aversion of 3.1.

tend to have lower equity exposure.

Lastly, we explore the dynamic factors driving heterogeneity in beliefs. There is a long literature documenting that investors extrapolate their beliefs across a number of settings.[6] We first find that investors extrapolate their beliefs from fund past returns. Taking advantage of turnover in 401(k) plan menus, we show that investors also extrapolate from past returns of new funds added to their plan menus, and so the extrapolation cannot be explained by inattention or inertia in rebalancing. Investors potentially form their beliefs based on the past returns reported in 401(k) plan brochures.

Next, we show how investors' personal experiences influence their beliefs. We find that local economic conditions, such as county-level GDP, population, and home price growth, are positively correlated with beliefs about market returns. For the subset of publicly traded employers, we also find that investors' expectations are positively correlated with the past performance of their employer, as measured by returns, investment, employment growth, and sales growth, even after controlling for industry-by-year fixed effects. This suggests that investors' expectations reflect their personal experience, which is consistent with evidence from other settings (e.g., Malmendier and Nagel, 2011, 2015).

In the cross section, beliefs are correlated with investor demographics and sector of employment. Over time, investors update their beliefs in response to local information such as GDP growth, business establishment growth, and population growth at the county level, *above and beyond* what is available from aggregate, macro-level information. Further, investors extrapolate future stock market performance based on the recent performance of their employer. These key findings—that expectations demonstrate systematic and predictable cross-sectional differences and, in the time series, are influenced by local factors—point to the importance of idiosyncratic experiences in the formation of beliefs. Our finding that investors respond to local information indicates that even potentially irrelevant information helps shape beliefs that have real stakes (Bordalo et al., 2022).

Given these factors driving the formation of beliefs, we explore the rationality of investor expectations, and, indirectly, address whether the belief formation strategies are rational. We find that investor forecast errors are predictable and consequently violate full information rational expectations, similar to the evidence documented in the literature.[7] Consistent with the evidence in Coibion and Gorodnichenko (2015) and Bordalo et al. (2018), we find that investor forecast revisions are negatively correlated with forecast errors, which suggests that investors overreact to news. One might expect that these patterns are driven by inexperience in financial markets. However, we find that beliefs of investors working in the financial sector

[6]For example, previous work documents extrapolation in the stock market (Benartzi, 2001; Greenwood and Shleifer, 2014), the housing market (Case et al., 2012), risk taking (Malmendier and Nagel, 2011), investment decisions (Gennaioli et al., 2016), and inflation markets (Malmendier and Nagel, 2015).

[7]For example, see Bacchetta et al. (2009); Coibion and Gorodnichenko (2012, 2015); Amromin and Sharpe (2014); Greenwood and Shleifer (2014); Gennaioli et al. (2016); Bordalo et al. (2019) among others.

are extrapolative, violate full information rational expectations, and tend to overreact to news.

The paper proceeds as follows: In Section 2, we describe the data used in our analysis. In Section 3, we present some basic facts about how portfolio allocations differ across investors and over time. We introduce our model and estimation procedure in Section 4. In Section 5 we present our baseline estimates and show how risk aversion and beliefs vary in the cross section. We explore the dynamic factors that explain the formation of investor expectations and test whether the expectations are rational in Section 6. Section 7 concludes.

Related Literature

There exists a rich theoretical literature on how households should invest[8] and how investment decisions vary with investor characteristics. We contribute to this literature by documenting how portfolio allocations vary across 401(k) plans. Consistent with the previous literature, we document that equity allocation is positively correlated with wealth (Heaton and Lucas, 2000; Wachter and Yogo, 2010; Bach et al., 2020; Fagereng et al., 2020) and education (Black et al., 2018); and is negatively correlated with age (Cocco et al., 2005) and minority status (Campbell, 2006). 401(k) plans provide a unique setting to study household portfolio choices because we observe the suite of investments available to participants and we can use variation in plan design/expenses to separately identify investor beliefs and risk aversion. Our empirical model provides insight into how differences in risk aversion and beliefs contribute to the patterns we see in the data.

To this end, our paper also relates to the literature using a revealed-preference approach to estimating beliefs and risk aversion across investors. Our methodology relates most closely to Egan et al. (forthcoming) who build and estimate a structural model to recover investor expectations and risk aversion using data from leveraged exchanged traded funds. We employ a related identification strategy, building on the work of Barseghyan et al. (2013), where we use exogenous variation in expenses to separately identify beliefs and risk aversion. Our framework is also related to that of Shumway et al. (2009), who use a revealed-preference approach to understand fund manager beliefs. A key distinction between our work and that of Shumway et al. (2009) is that we focus on separately identifying risk aversion and beliefs, and we focus on retail investors.[9] This type of revealed preference approach to asset pricing, by focusing on quantities rather than prices or returns, is also conceptually related to Berk and van Binsbergen (2016), Koijen and Yogo (2019a), and Heipertz et al. (2019). In their seminal work, Koijen and Yogo (2019a) develop a flexible characteristics-based demand system asset pricing model with

[8]This literature dates back to Markowitz (1952) and Merton (1969). See Campbell et al. (2002), Campbell (2006), Gomes et al. (2020), and Cochrane (2022) for a discussion of the literature.

[9]In contrast, Shumway et al. (2009) do not separately identify risk aversion and beliefs but instead recover beliefs of institutional investors up to an affine transformation that is scaled by risk aversion and translated by investors' shadow value of borrowing.

heterogeneous investors.[10] Our framework complements the work of Koijen and Yogo (2019a), taking a different approach to investor heterogeneity and leveraging a different identification strategy, which exploits variation in fund expense ratios, to address the endogeneity of demand and asset prices.

The two main methods used to study investor beliefs use either survey data or asset prices to measure beliefs. One strand of literature uses survey data to measure investor beliefs (Vissing-Jorgensen, 2003; Ben-David et al., 2013; Amromin and Sharpe, 2014; Greenwood and Shleifer, 2014 Nagel and Xu, 2019). While some researchers have been skeptical of surveys, recent evidence suggests they produce valuable information (Greenwood and Shleifer, 2014; Giglio et al., 2021). Our motivation is tied to the seminal results of Giglio et al. (2021), who document substantial and persistent heterogeneity in beliefs across retail investors. Using novel survey and account level data from Vanguard, Giglio et al. (2021) find evidence that beliefs are reflected in the portfolios of investors, especially when investors are attentive, are actively trading, and hold tax-advantaged accounts. When looking at tax-advantaged accounts, Giglio et al. (2021) estimate that a 1pp increase in beliefs about stock market returns is correlated with a 1.34-3.55pp increase in equity share, depending on the investor's characteristics (see column 7 of Table 4 in Giglio et al., 2021). Our baseline estimates imply a 1pp increase in beliefs about the stock market returns is correlated with a 3.68pp increase in equity share.[11] Despite using completely different samples and methodologies, we find estimates that are in line with Giglio et al. (2021), albeit slightly higher. Our approach provides insight into both investor risk aversion and beliefs when such survey data is unavailable.

The other strand of literature uses data on asset prices to recover investor beliefs (Ross, 2015).[12] While that literature uses data on asset prices to recover the distribution of beliefs of a single representative investor, we use data on allocations to recover the distribution of beliefs as well as risk aversion across investors. An advantage in our setting is that we observe plausibly exogenous variation in investment costs, which allows us to recover the distribution of both beliefs and risk aversion without making any assumptions about the structure of asset prices or beliefs.

Lastly, our paper relates to the literature on retirement savings (see Benartzi and Thaler, 2007 and Choi, 2015 for a discussion of the literature). A subset of this literature focuses on 401(k) enrollment and contributions and studies the effects of plan design such as automatic enrollment (e.g., Madrian and Shea (2001); Choi et al. (2007); Beshears et al. (2009); and

[10]The Koijen and Yogo (2019a) methodology has been extended to study other settings, including exchange rates (Koijen and Yogo, 2019b), cryptocurrencies (Benetton and Compiani, 2021), bonds (Bretscher et al., 2020), competition in the stock market (Haddad et al., 2021), and global equities (Koijen et al., 2019).

[11]The results in column (3) of Table 7 indicate that a 1 standard deviation increase in beliefs (2.3pp) is associated with a 8.56pp increase in equity share.

[12]Other recent examples include Jensen et al. (2019), Martin and Ross (2019), and d'Arienzo (2020). There is also a related strand of literature that focuses on robust identification of investor beliefs (Chen et al., 2020; Ghosh and Roussellet, 2020; Ghosh et al., 2020).

Carroll et al., 2009) and firm matching (e.g., Choi et al., 2002; Duflo et al., 2006; Dworak-Fisher, 2011). Due perhaps in part to the impact of this earlier literature, we find that plan participation is relatively high (74% on average) in our sample. While we find that participation is high conditional on eligibility, Yogo et al. (2021) documents that many households, especially low-income households, do not have access to employer-sponsored retirement plans and that providing access could increase retirement account participation by upwards of 10pp. Another strand of literature focuses on menu design and fees (Pool et al., 2016; Pool et al., 2020; Bhattacharya and Illanes, 2021). For example, Bhattacharya and Illanes (2021) develop a structural model of plan design and show how imperfect competition and agency frictions can lead to sub-optimal plan design. By contrast, we focus on the asset allocation decisions conditional on both participation and the 401(k) menu. Bekaert et al. (2017) document how both menu design and investor characteristics are related to international equity exposure in 401(k) plans. Previous work emphasizes the importance of behavioral frictions in 401(k) asset allocation decisions (Benartzi and Thaler, 2001, 2007). In a similar vein, we find that investor beliefs violate full information rational expectations and extrapolate from local experiences.[13] A growing body of research documents that such adaptive expectations could have significant implications for the macroeconomy and financial markets (Bordalo et al., 2018; Gennaioli and Shleifer, 2018; Bordalo et al., 2018; Malmendier et al., 2020).

2 Data

2.1 Sources

Our primary data set comes from BrightScope Beacon. BrightScope Beacon provides detailed plan and fund level information for ERISA defined contribution plans, covering 85% of plan assets. BrightScope collects the data either directly from plan sponsors, or from publicly available sources ranging from The United States Department of Labor (DOL) to the Securities and Exchange Commission (SEC). We focus on 401(k) defined contribution plans. The data set covers 70,000 different 401(k) plans over the period 2009-2019, resulting in roughly 450k plan-by-year observations. For each 401(k) plan, BrightScope reports annual data on the specific investment options available to participants and the total amount invested (across all plan participants) in each investment option. BrightScope does not provide individual investor level holdings data but provides holdings at the plan level. The data also includes details on the investment options in terms of the fee structure and type of funds. Because each 401(k) plan offers, on average, 26 different investment options, we have 11 million observations at the in-

[13]Previous research such as Bacchetta et al. (2009); Coibion and Gorodnichenko (2012, 2015); Amromin and Sharpe (2014); Greenwood and Shleifer (2014); Gennaioli et al. (2016); Bordalo et al. (2019) among others have found that beliefs violate full information rational expectations. See Malmendier (2021) for an overview of the literature on experience effects in finance.

vestment option-by-plan-by-year level, which is the unit of observation in our baseline analysis.

We merge our investment menu level data from BrightScope with additional data from the DOL Form 5500. The DOL Form 5500 data provides additional plan-by-year level details on plan participants, including the number of plan participants, the plan participation rate, employer contributions, and the share of participants that are retired.

We supplement our 401(k) data with mutual fund and stock return data from CRSP. CRSP provides daily level return data for stocks and open-end funds and quarterly level expense data for open-end funds. We merge the investment option-by-plan-by-year data in BrightScope with data from CRSP at the ticker-by-year level. Lastly, we include employment and demographic data at the county-by-industry-by-year level from the American Community Survey (ACS). We merge the ACS data with our 401(k) data based on the year, sponsor/employer industry (i.e., 2-digit NAICS), and the location of the employer's headquarters.

2.2 Summary Statistics

We start by documenting substantial heterogeneity in 401(k) holdings across plans and over time. We group investment options into six major asset classes: US equities, bonds, cash, target date funds, alternatives, and international equities. One minor complication in computing equity exposure is that some funds invest across asset classes (i.e., allocation funds such as Bridgewater's All Weather fund). When calculating US equity and bond shares for our reduced form analysis in Sections 2 and 3, we assume that non-target-date allocation funds hold sixty percent of their assets in US equities and forty percent in bonds. When estimating our quantitative model in Section 4, we calculate the equity/factor exposure of each fund using historical data.

Figure 1 displays the portfolio weights for six major asset classes across plan-by-year observations. The average plan holds 44% of the 401(k) assets in US equities,[14] but there is substantial heterogeneity across plans. The standard deviation of US equity allocations across plans is 19% with some plans having almost no money allocated to equities and others having 100% allocated to equities (Figure 1a).[15] Similarly, there is substantial heterogeneity in cash holdings across 401(k) plans. The average plan holds 11% in cash, but the standard deviation across plans is 13%.

Figure 2 displays the average portfolio weights for each of the major asset classes over time. In panels (a) and (b) we compute portfolio weights both excluding and including target date funds because target date funds tend to be the default option in most plans following the Pension Protection Act of 2006. There are some changes in the average holdings over time.

[14]Excluding multi-asset/allocation funds, the average is 39%.

[15]We find similar dispersion in equity exposures when we compute the equity beta for each portfolio (Appendix Figure A4). We also find similar patterns if we examine 401(k) plans that were created after the Pension Protection Act of 2006, which changed how 401(k) plans were designed (Appendix Figure A3).

Around the time of the financial crisis, investors increased the weight held in cash and bonds at the expense of US equities and international assets. Another key trend in the industry has been the rise of target date funds, consistent with the evidence documented in Parker et al. (2020).[16]

Table 1 displays the summary statistics for the BrightScope data. Panel 1a displays plan level summary statistics. The average plan has $85 million in assets and the average participant balance is $66 thousand. Employers accounted for 34% of all contributions with the remaining coming from plan participants. Participants, on average, can choose from 26 different investment options in the plan menu. The average plan has 1,261 participants.

The results also indicate that participation rates are quite high and that most eligible employees participate in 401(k) plans. At the median (mean) plan in our sample, 83% (74%) of eligible employees participate, which is consistent with estimates from the Survey of Consumer Finances.[17] The high participation rates are a relatively new phenomenon in the US. For example, in 1988, only 57% of eligible employees participated in 401(k) plans (Choi, 2015). Participation rates remained relatively high and constant over our sample period of 2009-2019.[18] The high participation rates may be a direct result of the earlier research, such as Madrian and Shea (2001), Choi et al. (2002), Choi et al. (2007), and Beshears et al. (2009), which emphasize how automatic enrollment increases 401(k) participation. While there has been concern about the lack of retirement savings in the US, these summary statistics suggest that the low retirement savings rates are driven by 401(k) plan eligibility rather than by 401(k) plan participation.

We also observe detailed information on each investment option. BrightScope Beacon classifies each investment option into eight different types of investment vehicles. The vast majority of investment options are structured either as mutual funds (61%) or as separate accounts (19%).

Table 1b displays investment option level summary statistics. BrightScope Beacon provides the latest expense ratios for each investment option, and we are able to obtain historical expense ratio data for those investment options structured as mutual funds using data from CRSP. As of 2019 the equal weighted average expense ratio was 57 bps.

Overall, the summary statistics presented in Figure 1 and Table 1 indicate that there is substantial variation in plan characteristics and holdings. Because our data is at the plan level rather than at the individual level, our summary statistics may understate the degree of heterogeneity in individual holdings. In the remainder of the paper, we explore the drivers of plan-level heterogeneity.

[16]Appendix Figure A1 shows a version where we do not attribute non target date allocation funds to US equity and bond assets. The sharp rise of allocation assets is mostly driven by target date funds, and the trends of equity and bond are similar when we do not consider non-target-date allocations.

[17]https://crr.bc.edu/wp-content/uploads/2020/10/IB_20-14.pdf

[18]Appendix Figure A2 displays participation and employer contribution rates over time. In Appendix Table A1, we examine how participation rates vary with the demographics of eligible participants. We find that participation is positively correlated with age and negatively correlated with minority status. However, consistent with the evidence in Yogo et al. (2021) we find no relationship between minority status and participation once we condition income and wealth.

3 How Do Asset Allocations Vary Across Investors?

The summary statistics presented in Section 2 indicate that there is substantial heterogeneity in asset allocations across plans. In this section, we explore what drives such heterogeneity. We find that asset allocations are highly correlated with participant demographics and employment. We also explore how much of the heterogeneity is driven by participant allocation decisions versus heterogeneity in 401(k) menus. We find that, while features of investment menus (such as the number of equity funds available) are correlated with investment decisions, differences in investment menus across plans do not explain the facts we document about the investment allocation decisions of investors.

We also examine other elements of the investment allocation process. We find evidence that investors appear to make at least partially informed decisions when selecting investment options. Investment decisions are sensitive to expense ratios and investors appear to rebalance their 401(k) portfolios over time. We also investigate whether investors use naive "1/N" diversification strategies by equally distributing their portfolio across all investment options. This type of behavior was first documented by Benartzi and Thaler (2001). Consistent with Benartzi and Thaler (2001), we find some evidence that investment allocations are correlated with the composition of the menu; however, a naive strategy only explains a small fraction of the variation in holdings and fails to explain the main patterns we observe in the data.

To understand how systematic differences in risk aversion and beliefs explain the facts documented here, we develop and estimate a structural model of portfolio choice. This analysis begins in Section 4.

3.1 Asset Allocation and Investor Characteristics

We examine how 401(k) portfolio allocation decisions vary across participant characteristics previously emphasized in both the theoretical and empirical literature. One unique feature of our setting is that all plan participants have access to all investment options available in the 401(k) menu. Thus, the patterns in allocation decisions we document are not driven by participation costs, which has been emphasized in other settings.[19]

Demographic Characteristics: We examine how investment allocations vary across investor demographics in the following regression:

$$Share\, in\, US\, Equities_{mt} = X_{mt}\beta + \epsilon_{mt}. \tag{1}$$

Observations are at the 401(k) plan-by-year level. The dependent variable $Share\, in\, US\, Equities_{mt}$ reflects the share of assets held in equities in plan m at time t. When computing the share of

[19] See Campbell (2006) and Gomes et al. (2020) for a discussion of the literature.

assets held in US equities we exclude target date funds because they tend to be the default option in 401(k) plans.[20]

We consider demographics, industry, and plan variables in X_{kt}. Following the literature, we focus on age, income, housing wealth, and race using county-by-industry-by-year level demographics information from the ACS. Since we do not perfectly observe participant demographics, this may introduce measurement error in our demographic covariates and could attenuate some of our results.[21] We also include several plan-level characteristics using Form 5500 data. The Form 5500 data includes plan-by-year level information on the average account balance of plan participants, the share of participants that are retired, and plan age.

We present the corresponding estimates in Table 2. We include time, county, and industry fixed effects in each specification. Columns (1)-(11) display univariate regressions, and the specification reported in column (12) includes the full set of controls. For ease of interpretation the independent variables are in units of standard deviation.

Overall, we find significant relationships between investor characteristics and equity allocation. Plans with wealthier participants, measured by average account balances, allocate more towards equities. The results in column (12) indicate that a one standard deviation increase in the average account balance is correlated with a 0.89 pp increase in equity exposure. Likewise, more educated households have higher equity allocation. A one standard deviation increase in the share of college educated individuals is correlated with a 0.86 pp increase in equity allocation. One the other hand, age, the share of retired participants and the share of minorities (Hispanic or black) are negatively correlated with equity exposure. One standard deviation increases in participant age and the share of retired participants are associated with a 0.17 and 0.40 pp decline in US equity holdings, respectively (column 12).

In Appendix A.1, we provide additional details of these relationships, including connections to previous findings in the literature and potential explanations, and we examine allocations to other asset classes such as bonds and international equities.

Employment: Figure 3 displays the distribution of equity exposure by the 2-digit NAICS of the employer. Median equity exposure varies across sectors, ranging from 53.1 percent in Public Administration to 62.6 percent in Information. Such variation could potentially be consistent with background risk, such as shocks to labor income.[22] However, the pattern across sectors suggests that risk is not the only factor driving allocation decisions. For example, it is not

[20]As mentioned in Section 2.2, we assume that non-target-date allocation funds hold sixty percent of their assets in US equities. Our main findings are robust under other assumptions such as if we include target date funds, exclude all allocation funds, etc.

[21]As a robustness check, we replicate our findings for smaller firms (below median) where measurement error is likely smaller in the Appendix.

[22]Households with higher undiversified labor risks may effectively be more risk averse and should invest more cautiously (Heaton and Lucas, 2000; Viceira, 2001). In addition, in sectors with more flexible labor conditions, households can adjust labor supply in response to investment returns, and thus increase willingness to take financial risk (Bodie et al., 1992, Farhi and Panageas, 2007).

obvious that employment in the Public Administration sector would be substantially riskier than employment in the Information sector. Instead, some of the differences across sectors may be explained by differences in risk aversion and beliefs, in addition to underlying risk. Our results in Figure 3 are probably best explained by a mixture of these factors. We delve into these sector differences in equity exposure further in Section 5 to understand if they can be explained by differences in beliefs and/or risk aversion.

3.2 Understanding How Investors Form Portfolios

There exists a long theoretical literature illustrating how rational investors should form portfolios (e.g., Merton, 1969), yet the empirical literature documents that portfolio theory often fails to match how households invest in practice (Benartzi and Thaler, 2007; Cochrane, 2022). Here, we explore which factors appear to drive investor portfolio decisions. Based on the previous literature, we focus on how investors form portfolios based on expenses (Hortaçsu and Syverson, 2004) and the composition of the menu (Benartzi and Thaler, 2001). We also explore the rebalancing behavior of investors.

Our results suggest that investor decisions appear at least partially informed and attentive: investors are sensitive to fees and appear to rebalance their portfolios over time. While we find evidence suggesting that the composition of the 401(k) menu is correlated with investment decisions, the evidence is weaker than what has been documented previously in the literature.[23] These facts are important for motivating the empirical model that we build and estimate in Sections 4 and 5. In particular, an investor's sensitivity to fees is a key moment for separately identifying risk aversion from beliefs in our quantitative model.

3.2.1 Responding to Fees

In any portfolio choice model, investors trade off risk with expected returns. Measuring how investors respond to exogenous changes in fund expense ratios provides insight into this trade-off as expense ratios directly impact the expected returns of the fund. This also provides some insight into the optimality of an investor's investment decisions.

We start with a simple cut of the data by looking at the equal-weighted distribution of fund expenses relative to the asset weighted distribution of fund expenses in Figure 4. Panels (a) and (b) show a stark contrast between the equal weighted and asset weighted distributions of expenses. The asset weighted distribution is shifted dramatically to the left relative to the equal weighted distribution. The average fund appearing on an investor's 401(k) menu charges an

[23]Prior work suggests that some investors may follow naive diversification strategies by allocating their portfolio equally across the funds in their 401(k) menu, which suggests that the composition of the 401(k) menu has a large impact on investment allocations. We focus on naive diversification in Section 3.2.2. In Appendix Table A3, we consider a richer set of plan menu features and show that menus are largely uncorrelated with participant demographics

expense ratio of 57 bps; however, the average expense ratio paid by investors is 26 bps. This is driven in part because investors tend to tilt their portfolio allocations towards inexpensive funds and provides prima facie evidence that investor demand is sensitive to fees.

We examine the relationship more formally in the following demand specification:

$$\ln Share_{kmt} = \alpha p_{kt} + \phi_{mt} + \phi_{\tau(k)t} + \xi_{kmt}. \tag{2}$$

Observations are at the fund-by-plan-by-year level where we exclude target date funds. The dependent variable $Share_{kmt}$ measures the share of assets held in fund k in plan m at time t relative to the total assets in plan m at time t. Fund expense p_{kt} is the independent variable of interest. We include plan-by-year (ϕ_{mt}) and fund-type-by-year ($\phi_{\tau(k)t}$) fixed effects. We define fund type $\tau(k)$ based on the fund's classification in both Morningstar and BrightScope and whether it is a index/passive fund (i.e., Morningstar Category x BrightScope Category x Passive). Including plan-by-year fixed effects is important because it allows us to measure how investors trade off relative differences in expenses among the funds available in the investor's 401(k) menu rather than differences across 401(k) menus, which may be correlated with plan size. While we present eq. (2) as a simple linear specification, by including plan–by-year fixed effects, eq. (2) directly corresponds to the workhorse discrete-choice demand model developed in Berry (1994) that is commonly used in the industrial organization literature.[24]

One concern with estimating demand is that fund expenses are potentially endogenous. For example, if investors are particularly optimistic about a fund (e.g., high ξ_{kmt}) the fund provider may find it optimal to increase the fund expense ratio. This type of endogeneity would typically bias our estimate of α upwards (i.e., α is less negative) such that investors appear less sensitive than they actually are. To account for the potential endogeneity, we instrument for fees using Hausman-type instruments. Specifically, we use the average fee charged by the mutual fund provider in other Lipper objective investment categories in the same year.[25]

We report our demand estimates in Table 3. Column (1) displays the OLS results and column (2) displays the corresponding IV results. Note that OLS and IV estimates are quite

[24]Following the setup in (Berry, 1994) , the market share of product k in market m can be written in logs as

$$\ln share_{kmt} = \alpha p_{kt} + \xi_{kmt} - \ln\left(\sum_{k' \in \mathcal{K}_{mt}} \exp(\alpha p_{k't} + \xi_{k't})\right),$$

where ξ_{kmt} captures unobserved product characteristics and $\mathcal{K}_{mt}$ is the set of available products available in market m at time t. In the context of 401(k) choice, k refers to the fund and markets are defined based on the 401(k) plan menu. The plan-by-year fixed effect in eq. 2 absorbs the non-linear term $\ln\left(\sum_{k' \in \mathcal{K}_{mt}} \exp(\alpha p_{k't} + \xi_{k't})\right)$ which is constant within a plan-year. This type of demand system has been used in a number of other financial applications such as demand for bank deposits (Dick, 2008; Egan, Hortaçsu, and Matvos, 2017; Wang, Whited, Wu, and Xiao, 2018; Egan, Lewellen, and Sunderam, 2022), bonds (Egan, 2019), credit default swaps (Du et al., 2019), insurance (Koijen and Yogo, 2016, 2022), mortgages (Benetton and Compiani, 2021) and investments more generally (Koijen and Yogo, 2019a,b; Koijen et al., 2019).

[25]See section 4.3.2 for further discussion of endogeneity of fees and the instruments.

similar, so the potential endogeneity concern appears minimal. The results indicate that, as expected, investors are sensitive to expenses. The results in column (2) indicate that a 10 bp increase in fees is associated with a 6.7% decrease in demand. In the context of the discrete choice demand system developed in Berry (1994), the estimates in column (2) correspond to a demand elasticity of -0.40.[26] We find that demand is relatively inelastic, which is consistent with a recent literature highlighting potentially inelastic demand in the stock market (Koijen and Yogo, 2019a). Collectively, the results suggest that investors respond to expenses, which is perhaps not surprising given that expenses are often a salient feature when investors make 401(k) decisions.

3.2.2 Comparing Allocations to Naive Diversification

Previous results in the literature have suggested that some investors follow naive diversification strategies in which they simply split their allocation evenly across all of the options in their retirement plans (Benartzi and Thaler, 2001). Investor behavior along these lines would enable employers to adjust the share of equity allocation by simply increasing the relative number of equity options in the retirement plan.

Using our data, we consider an implication of these naive diversification strategies for aggregate investment trends. Holding fixed the balance in each plan in each year, we simulate the counterfactual holdings if all investors simply allocated their funds evenly across all funds in the menu and, alternatively, evenly across the eight categories of funds in our data (e.g., Bond Funds, Cash / Stable Value, International Stock, US Large Cap Stock, etc.).

Figure 5(a) shows the allocation to US equities in our data and the counterfactual "1/N" naive investment strategies. From 2009 through 2019, excluding target date funds, there has been a steady increase in the share of retirement assets allocated to US equities, as shown by the solid line. Conversely, if investors were simply allocating funds evenly across funds (dashed line) or fund categories (dash-dotted line), US equities would have declined as a share of assets. Over time, US equity funds are making up a smaller share of fund choices, even as they constitute a greater share of retirement assets. In Figure 5(b) and (c), we compare allocations to US equities with the predicted allocations based on the 1/N strategy. Panel (b) displays the observed distribution of equity allocations with those predicted by the 1/N strategy. The results suggest that the observed dispersion in equity allocations is much more disperse than what would be implied if all investors used a 1/N strategy. Furthermore, panel (c) indicates that the

[26] We compute the demand elasticity assuming a market share of 1/26 and fee of 0.61pp. It is also useful to compare our estimates with those found in other financial markets. For example, recent studies have found that demand is inelastic in bank deposit markets (0.20-0.75; Dick (2008); Egan et al. (2017); Xiao (2020); and Egan et al. (2022)), privatized social security markets (0.3-1; Hastings et al. (2017)), and equity brokerage markets (0.47; Di Maggio et al. (2021)). In contrast, other researchers have found that demand is more elastic in life insurance markets (2.18; Koijen and Yogo (2016)) and mortgage markets (2-6; Buchak et al. (2018); Robles-Garcia (2019); Benetton (2021)). This is intuitive and these results suggest that within a 401(k) plan, the available funds are less substitutable than are mortgage providers and life insurers.

1/N strategy explains very little of the variation in equity allocations.

We also replicate the baseline analysis presented in Benartzi and Thaler (2001). Appendix Table A6 shows that in our setting the menu composition is correlated with investment decisions, though we find a weaker relationship. Using the same specification, we obtain R^2 ranging from 0.03 to 0.18 whereas R^2 ranges from 0.25 to 0.62 in Benartzi and Thaler (2001). This weaker relationship with naive investment strategies is consistent with the findings in Huberman and Jiang (2006), where "the available fund mix and number of funds offered hardly explains participants' choices of funds." Part of this may be due to the sample composition. Benartzi and Thaler (2001) study a cross-section of 170 plans in 1996 where the average plan has 6.8 different investment options. We study a much larger and more recent sample where the average plan has 26 options. Investors may have become more sophisticated in the past twenty years, and it is possible that investor behavior changes when facing a menu with more options, holding fixed investor sophistication.

3.2.3 Rebalancing Behavior

Lastly, we examine investor rebalancing behavior. Figure 2 displays average holdings over time. Investors' equity exposure is slightly decreasing over this period. Given that the S&P 500 Index increased almost 200% during our sample period, the fact that the investors' shares in US equities did not increase dramatically suggests that investors must be rebalancing their portfolios over this period. Survey evidence shows that in 2020 (2009) roughly 17% (15%) of DC participants changed the asset allocation of their account balance and 10% (19%) changed the asset allocation of their contribution.[27]

To understand investor's rebalancing behaviors more systematically, we calculate the autocorrelation in plan holdings in Table 4. Specifically, we calculate the variable $Expected\ Portfolio\ Weight_{ijt}$ which assumes that the portfolio weight of fund i grows by the return of fund i relative to the total return of the 401(k) portfolio over the same period. The results in column (1) indicate that the correlation between $Expected\,Portfolio\,Weight_{ijt}$ and $Portfolio\,Weight_{jt}$ is 0.77. Part of the reason investors rebalance is because 401(k) menus turnover quite frequently as documented in Sialm et al. (2015). If we restrict our attention to those plans that have been outstanding for at least one year, roughly 20% of 401(k) investment options were not available in the previous year. If the investment option is not available in the previous year, $Expected\,Portfolio\,Weight_{ijt}$ is zero by construction. Turnover among 401(k) menus and providers induces participants to rebalance their portfolios. In column (2), we replicate our analysis where we restrict our attention to only those investment options that were available in the previous year. Not surprisingly, the autocorrelation in holdings is higher if we exclude changes in the investment menu. Overall, the results suggest that, while there

[27] See https://www.ici.org/system/files/2021-09/21_rpt_recsurveyq2.pdf. ICI reports rebalancing activity for the first half of 2009 and 2020, which we annualize by multiplying them by two.

is persistence in portfolios, there is also variation in investor portfolios over time. Since fund fees are relatively persistent in the data and investors' beliefs might also be persistent, one might naturally expect there to be a large amount of persistence in portfolios over time even if investors are actively rebalancing their portfolios.

Overall, our results suggest that investors' portfolio allocations appear to be at least partially informed, and so their decisions reveal information about their beliefs and risk aversion. In the next section, we develop and estimate a model of portfolio choice that allows us to interpret the facts documented in Sections 3.1 and 3.2 in terms of differences in investor beliefs and risk aversion.

4 Model

Motivated by the above findings, we model each investor's 401(k) portfolio allocation as a mean-variance decision problem. Each investor trades off her subjective and potentially biased expectation of the return of investing an additional dollar in one of the available 401(k) investment options with the additional risk scaled by risk aversion. Using this framework, we show how to separately identify an investor's beliefs about the expected returns of each asset and risk aversion.

We also use our estimates of beliefs and risk aversion to better understand the portfolio allocations of investors. Without our structural framework, an analysis of portfolio allocations provides limited insight into investors' decisions. That is because portfolio allocations are a function of both 401(k) plan design and investor preferences/beliefs. For example, if we were to observe an investor with a relatively small equity allocation it could be because: (i) the investor is risk averse, (ii) the investor is pessimistic about the return of the market, and/or (iii) the equity investment options in the investor's 401(k) plan are expensive. Unlike portfolio allocations, our estimates of beliefs and risk aversion adjust for the menu of funds available in each investor's 401(k) plan. If two identical sets of investors faced different plans menus, they may have different portfolio allocations. With sufficient variation in funds within a menu, our methodology would recover the same set of beliefs despite the different observed allocations.

4.1 Investor's Problem

Each investor i must form portfolios from the set of securities $k = 1, ...K_i$ and a risk-free asset. We assume investors have mean-variance preferences with risk aversion λ. Investors choose the $K \times 1$ vector of weights $\boldsymbol{\omega}_i$ to maximize

$$\max_{\omega} \boldsymbol{\omega}_i'(\boldsymbol{\mu}_i - \boldsymbol{p}) + (1 - \boldsymbol{\omega}_i'\mathbf{1})R_F - \frac{\lambda}{2}\boldsymbol{\omega}_i'\Sigma\boldsymbol{\omega}_i,$$

where $\boldsymbol{\mu}_i$ is a vector of investor i's expectations of fund returns, $\boldsymbol{p}$ is a vector of fund expenses, R_F is the risk-free return, Σ is the $K \times K$ covariance matrix of expected fund returns, and λ_i is risk aversion. The corresponding first order condition is

$$\boldsymbol{\mu}_i - \boldsymbol{p} - \mathbf{1}R_F = \lambda_i \Sigma \boldsymbol{\omega}_i.$$

We then have $\kappa_i \in \{1, ..., K_i\}$ first order conditions for every investor.

4.2 Empirical Framework

We assume that the return of each asset k follows a factor structure with L orthogonal factors f_{lt} and idiosyncratic component ϵ_{kt}. By construction the factors and idiosyncratic component each have a variance of one. We can then write returns as:

$$R_{kt} = \sum_{l=1}^{L} b_{klt} f_{lt} + \sigma_{kt}\epsilon_{kt},$$

yielding a covariance matrix

$$\Sigma_t = \boldsymbol{b}_t I_L \boldsymbol{b}_t' + \boldsymbol{\sigma_t} I_K \boldsymbol{\sigma_t}'.$$

The factors are orthogonal by construction. We assume that the idiosyncratic component is uncorrelated across securities.

We assume investors agree on the factor structure and the loadings $(\boldsymbol{b}_t, \boldsymbol{\sigma}_t)$. Thus, differences in beliefs about returns for an asset k arise from differences in expected realizations of factors and the idiosyncratic component, $\mu_{ikt} = E_i[R_{kt}] = \sum_{l=1}^{L} b_{klt} E_i[f_{lt}] + \sigma_{kt} E_i[\epsilon_{kt}]$.

We can then rewrite the above first order condition for each security k as

$$\mu_{ikt} - p_{kt} - R_F = \lambda_{it} \left(\sum_{l=1}^{L} b_{klt} \left(\sum_{j=1}^{K} b_{jlt}\omega_{ijt} \right) + \sigma_k^2 \omega_{ikt} \right). \tag{3}$$

The term on the left hand side reflects the expected return net of fees of investing an additional dollar in fund k, and the term on the right hand side reflects the additional risk of investing an additional dollar in security k.

In the data, we do not observe each investor i's portfolio but instead observe the aggregated portfolio for all investors participating in the same defined contribution retirement plan m. Let $\mathcal{I}_m$ denote the set of individuals participating in defined contribution plan m and A_i denote investor $i's$ total portfolio value. We can then write the value-weighted average of the first order conditions (eq. 3) across all individuals participating in defined contribution plan m as

$$\left(\frac{1}{\sum_{i \in \mathcal{I}_m} A_i} \right) \sum_{i \in \mathcal{I}_m} A_i \left(\mu_{ikt} - p_{kt} - R_F \right) = \lambda_{mt} \left(\frac{1}{\sum_{i \in \mathcal{I}_m} A_i} \right) \sum_{i \in \mathcal{I}_m} A_i \left(\sum_{l=1}^{L} b_{klt} \left(\sum_{j=1}^{K} b_{jlt}\omega_{ijt} \right) + \sigma_k^2 \omega_{ikt} \right),$$

where we assume that all investors in plan m have the same risk aversion λ_{mt}. This simplifies to

$$\bar{\mu}_{kt}^{(m)} - p_{kt} - R_F = \lambda_{mt}\left(\sum_{l=1}^{L} b_{klt}\left(\sum_{j=1}^{K} b_{jlt}\bar{\omega}_{jt}^{(m)}\right) + \sigma_k^2\bar{\omega}_{kt}^{(m)}\right),$$

where $\bar{\mu}_{kt}^{(m)}$ is the average (dollar-weighted) expected return of asset k at time t across investors participating in defined contribution plan m that purchase asset k. The weight $\bar{\omega}_{kt}^{(m)}$ is the average (dollar-weighted) portfolio weight.

Given the factor structure $\boldsymbol{b}_t$ and the idiosyncratic variance $\boldsymbol{\sigma}_t$, we can compute the risk associated with each fund k. We can then estimate the linear regression equation:

$$\left(\sum_{l=1}^{L} b_{klt}\left(\sum_{j=1}^{K} b_{jlt}\bar{\omega}_{jt}^{(m)}\right) + \sigma_k^2\bar{\omega}_{kt}^{(m)}\right) = \theta_{mt}p_{kt} + \epsilon_{kt}, \quad (4)$$

where the parameter θ is the negative inverse of risk aversion (i.e., $\theta = \frac{-1}{\lambda}$) and ϵ_{kt} is equal to average investor beliefs divided by risk aversion (i.e., $\epsilon_{kt} = (\bar{\mu}_{kt}^{(m)} - R_F)/\lambda_{mt}$). Eq. (4) is the heart of our estimation strategy. Identification requires exogenous variation in the fees investors pay for each investment option that is orthogonal to average investor beliefs (ϵ_{kt}). Given exogenous variation in fees, we are able to recover the parameter θ_{mt} and consequently risk aversion λ_{mt}. In principle, with a sufficient number of funds per plan, we could nonparametrically identify separate values of risk aversion for each plan and year combination. Given risk aversion, we can recover average beliefs as $\lambda_{mt}\epsilon_{kt} = \bar{\mu}_{kt}^{(m)} - R_F$.

4.3 Implementation

4.3.1 Risk

To estimate risk aversion and recover investor beliefs, we need to estimate the factor structure of fund returns ($\boldsymbol{b}_t, \boldsymbol{\sigma}_t$). We estimate the factor structure using a 6-factor model where we include the Fama French 3 factors and three bond factors: the excess return of long term government bonds; the excess return of investment-grade bonds, the excess return of high-yield bonds.[28]

We estimate factor loadings for each mutual fund and equity in CRSP using weekly return data over the previous ten years where we allow factor loadings to vary year-to-year. We then merge the estimated factor loadings with our BrightScope data at the fund-by-year level using mutual fund and stock tickers. Our data also contains non-mutual fund and stock options, such as separate accounts. For these investment options, we do not observe high-frequency

[28]We calculate long term government bonds returns using Vanguard's Long-Term Treasury Fund (VUSTX) returns, the investment grade bond returns using Vanguard's Long-Term Investment-Grade Fund (VWESX) returns, and high yield bond returns using Vanguard High-Yield Corporate Fund (VWEHX) returns. We calculate excess returns relative to the risk free rate as reported in the Fama and French database.

data, but we do observe their category classifications. We calculate the risk associated with these investment options based on the average risk of all other funds that belong to the same Morningstar category in the same year.[29]

As a robustness check, we also consider a simpler factor structure where we construct the factors by forming equal weighted portfolios based on the broad BrightScope categories reported in Table 1a, with the idea that investors think of risk in terms of broad asset classes (e.g., bonds, international stocks, cash, etc.). We also estimate a 55-factor model following Shumway et al. (2009). Estimates of beliefs and risk aversion using these alternative methodologies are highly correlated with our baseline estimates. We provide comparison statistics in Table A7. In Appendix A.3, we also explore the case where investors account for labor income risk and find that investors behave as if they neglect labor income related risks.

4.3.2 Expenses

We determine fund expenses using data from CRSP. One concern is that fund fees may be endogenously related to investor beliefs. For example, if a mutual fund provider anticipated that investors were optimistic about the returns of a particular fund, the fund provider might find it optimal to increase its expense ratio. This endogeneity would result in an upward bias in θ in eq. (4).

To help address this concern, we include plan-by-year fixed effects and fund classification-by-year[30] fixed effects in our main specification. Thus, we allow fees to rise endogenously in response to expectations of investors in specific plans or for specific fund categories in specific years, and we identify model parameters based on variation in expenses within plan-by-year and within classification-by-year. After including these fixed effects, the potential endogeneity concern would then be that, conditional on a 401(k) plan and fund classification, the residual variation in expenses is correlated with the residual variation in investor beliefs for specific funds. For example, suppose that (i) Fidelity anticipates that participants in IBM's 401(k) plan have more optimistic beliefs about Fidelity's Large Cap Growth Index Fund relative to the other investment options in IBM's 401(k) plan (average absorbed by plan-by-year fixed effects) and relative to average beliefs about other large cap growth funds (average absorbed by classification-by-year fixed effects) and, as a result, (ii) Fidelity increases the expense ratio it charges on its Large Cap Growth Index Fund. While certainly possible, the fact that mutual fund fees are infrequently updated and set uniformly helps alleviate these endogeneity concerns.

Nevertheless, to account for the potential endogeneity of fees, we instrument for fees using Hausman-type instruments as used in Section 3.2.1. Specifically, we use the average fee charged by the same mutual fund provider in other Lipper investment objective categories.[31]

[29]For a handful of options we do not observe the Morningstar category. For these funds we calculate risk based on the average risk of all other funds that belong to the same BrightScope category in the same year.

[30]Fund classification categories include, e.g., US Equity Large Cap Value Equity, Real Estate Equity, etc.

[31]When forming the instrument for fund k in plan m, we exclude all funds appearing on the menu for plan m

This instrument will be relevant (correlated with fees) when a provider's cost of operating a mutual fund is correlated with its costs of operating its other mutual funds, perhaps as a result of the provider's scale and technology. The instrument meets the exclusion restriction (provides exogenous variation) when participants' beliefs about the idiosyncratic expected returns of a given fund (after controlling for plan-by-year and category-by-year fixed effects) are, on average, uncorrelated with fees a provider charges on its funds from different Lipper investment objective categories. We consider both of these conditions to be plausible in our setting. A threat to exogeneity would be that, for example, an investor's belief about the expected return of Fidelity's Large Cap Growth Index Fund is correlated with the expenses Fidelity charges on its bond funds.

4.3.3 Portfolio Weights

We construct portfolio weights using total assets (across all participants in the plan) for each investment option and year reported in BrightScope. When constructing portfolio weights we treat all investment options categorized in BrightScope as "Cash/Stable Value" as risk-free assets. We also exclude funds classified in BrightScope as target date funds because these funds are often the default option and tend to be held by passive investors. However, as reported in Appendix Table A7, we find qualitatively similar estimates if we include target date funds in our analysis.

4.4 Estimation

We estimate the empirical analog of the investor's first order conditions for choosing an optimal portfolio (eq. 4) in the following regression specification:

$$\varsigma^2_{mkt} = \theta_{mt} p_{mkt} + \phi_{mt} + \phi_{j(k)t} + \epsilon_{mkt}, \tag{5}$$

where

$$\varsigma^2_{mkt} = \left(\sum_{l=1}^{L} b_{klt} \left(\sum_{j=1}^{K_i} b_{jlt} \bar{\omega}^{(m)}_{jt} \right) + \sigma^2_k \bar{\omega}^{(m)}_{kt} \right)$$

and ϕ_{mt} and $\phi_{j(k)t}$ are plan-by-year and fund type-by-year fixed effects. Here, subscript m denotes specific 401(k) plans, and $j(k)$ denotes fund type based on the fund's classification in both Morningstar and BrightScope as well as whether the fund is an index/passive fund. Thus, the fixed effect $\phi_{j(k)t}$ is a quadruple interaction term (i.e., Morningstar Category $\times$ BrightScope Category $\times$ Passive $\times$ Year). Observations are at the investment option-by-plan-by-year level. Because each observation reflects the average behavior of plan participants, we weight each observation by the total assets of the 401(k) plan when estimating eq. (5). Our estimates allow

when calculating the average fee charged by the mutual provider who manages fund k.

us to recover risk aversion, as $\widehat{\lambda}_{mt} = -\frac{1}{\hat{\theta}_{mt}}$. In principle, risk aversion is nonparametrically identified for each plan-year, provided a sufficient number of funds per plan. In practice, we parameterize θ_{mt} to allow for some flexibility.

Our empirical framework also allows us to recover the average expected returns within investors in a 401(k) plan for each investment option available in the plan. We recover the average beliefs for each investment option based on our estimate of θ_{mt}, our estimated fixed effects, and the residual from eq. (5):

$$\widehat{\bar{\mu}_{kt}^{(m)} - R_F} = -\frac{1}{\hat{\theta}_{mt}} \left(\hat{\phi}_{mt} + \hat{\phi}_{j(k)t} + \hat{\epsilon}_{mkt} \right). \tag{6}$$

Given each investor's beliefs about the expected return and the factor loadings for each investment option/fund, we can use the estimated distribution of beliefs to recover investors' expectations of the market return. We estimate the plan-by-year average expected market return at time t for each plan m based on the regression:

$$\widehat{\bar{\mu}_{kt}^{(m)} - R_F} = \delta_{mt} b_{1kt} + \eta_{mkt}, \tag{7}$$

where b_{1kt} is the loading for fund k on the market factor at time t. Observations are at the fund-by-plan-by-year level. The parameter δ_{mt}, which varies at the plan-by-year level, reflects the average expected return of the market across participants in plan m at time t. Note that because the other factors are orthogonal to the market by construction, we do not need to control for the other factors in eq. (7).

4.5 Identification and Interpretation

We estimate risk aversion by measuring how investors trade off risk and expected returns in eq. (5). Specifically, we estimate risk aversion by examining how investors adjust their portfolio risk exposure in response to changes in expense ratios. Changes in expense ratios are equivalent to shifts in expected returns, allowing us to calculate risk-return tradeoffs. An investor optimally sets the expected return of an investment equal to the marginal risk, scaled by risk aversion. Our approach relies on the following assumptions: we can correctly measure investor's beliefs about risk; investors make allocation decisions considering their retirement accounts only; investors solve a myopic portfolio problem; and investors only trade off risk and expected returns. We discuss how the interpretation of our estimates of beliefs and risk aversion would change if our baseline assumptions are violated. It is important to emphasize that we do not impose rational beliefs in our analysis and our framework allows for behavioral biases and mistakes in investor beliefs.

Measurement Error in Risk: We assume that investors understand and agree on the risk of their portfolios, and that we, as the econometrician, assess risk in the same way. Investors may have heterogeneous beliefs about risk or use different models for assessing risk, both of which could introduce measurement error into the dependent variable ς^2_{mkt}. Provided that the measurement error is orthogonal to our Hausman instrument, our instrumental variable estimate of risk aversion will still be consistent. While this does not impact our measurement of risk aversion, it will impact the beliefs we recover in the data. Rather than recovering beliefs $\bar{\mu}^{(m)}_{kt} - R_F$, we will recover beliefs plus the measurement error in risk. Since we average investor portfolios at the plan level, any measurement error that is mean zero and is uncorrelated across investors within a plan will not affect our results. However, if the measurement error is not mean zero or is correlated across investors within a plan, the estimates of beliefs we recover may be biased.

Outside Assets: By focusing on 401k investments, we only observe part of an investor's overall portfolio. According to 2019 Survey of Consumer Finance (SCF), the average individual has 31% of her financial assets in retirement accounts, compared to 7% in other investment funds, directly owned stocks, or bonds.[32] On average, individuals hold 57% in cash (deposit, money market, etc.). Thus, retirement accounts represent the vast majority of risky financial assets for many individuals. In addition, human capital measured by net present value of future income approximates the risk and return profile of bonds for most individuals, and hence constitutes another important financial asset. Although retirement accounts are the primary source of risky equity for most investors, we could potentially over-estimate an investor's equity share by ignoring outside cash holding and human capital.

To see how this would impact our estimates of beliefs and risk aversion, suppose that the true measure of risk is $\varsigma^2_{mkt} = h_{mt}\tilde{\varsigma}^2_{mkt} + \nu_{mt}$, where h_{mt} is the fraction of assets held in retirement accounts and ν_{mt} captures the additional risk from non-retirement account holdings. Rather than recover the true risk aversion, we will recover risk aversion scaled by h_{mt}.[33] With additional information on h_{mt}, we can adjust our risk aversion accordingly. For example, given that investors on average hold 31.4% in retirement accounts, we can divide our estimated risk aversion of 3.55 by 0.314 to obtain the true risk aversion of 11.3. The share h_{mt} likely varies across individuals, especially when we consider human capital. This could help explain why, as we discuss below, older investors behave as if they have higher risk aversion.

Our estimates of investor beliefs will capture an investor's true beliefs plus the term coming

[32] Retirement accounts and cash make up 36% and 16% of total dollar amount of financial assets. Pooled investment funds and directly owned stock represent 22% and 15% of total financial assets, but they are concentrated among few individuals, and so the average fractions in these categories are much smaller.

[33] Our estimate $\hat{\theta} = \frac{Cov(\varsigma^2/h - \nu/h, p)}{Var(p)} = \theta/h$, and so $\hat{\lambda} = h\lambda$. Since h_{mt} and ν_{mt} only vary at plan-by-year level, their variation is absorbed by plan-by-year fixed effects and does not cause endogeneity issues.

from additional risky asset holdings ν_{mt}/θ.[34] Since $\nu > 0, \theta < 0$, this will cause our estimates of beliefs to be biased downwards if investors have other equity assets outside of their 401k. Since the SCF data shows that 401(k) accounts are the primary source of risky assets for most households, this suggests that associated bias may be small.

Dynamic Allocation Across Multiple Periods: We model an investor's allocation decision as a myopic portfolio choice problem. It is well known that when investors have power utility and return distributions are independent over time, long-term portfolio choice is equivalent to myopic portfolio choice. More general time-varying returns could introduce intertemporal hedging demand, which would not impact our estimates of risk aversion but would potentially impact and be captured in our estimates of beliefs.[35] For example, if equity is mean reverting, lower unexpected return today is correlated with better investment opportunities in the future. This would result in a positive hedging term and would potentially bias our estimates of beliefs upwards.

Optimization Error and Inattention: We assume that investors actively trade-off and equate marginal risk and return when making investment decisions. There are a few reasons this could be violated in the data. Suppose that marginal risk is equal to expected returns plus some vector of optimization errors $\boldsymbol{\zeta_i}$:

$$\lambda\Sigma\boldsymbol{\omega}_i = \boldsymbol{\mu}_i - \boldsymbol{p} - \mathbf{1}R_F + \boldsymbol{\zeta_i}. \tag{8}$$

One could interpret this optimization error as either a true error term or it could be capturing unobserved preferences of consumers. For example, it could be the case that even conditional on the risk and expected returns of a fund, investors have preferences for one fund over another. This type of optimization error would impact our estimation in the exact same way as a noisy measure of risk; if the optimization error is either not mean zero and/or correlated across investors within a plan, the risk aversion estimate would still be consistent but the beliefs estimates would reflect this preference (and potentially be biased)..

Relatedly, one might be concerned that investors do not actively trade off expected returns with risk. For example, investors may be inattentive such that only a fraction of investors

[34]To see this, we express our recovered belief as follows, where we omit subscripts:

$$\widehat{\bar{\mu} - R_F} = -\frac{\hat{\epsilon}}{\hat{\theta}} = -\frac{(\frac{\varsigma^2-\nu}{h} - \frac{\theta}{h}p)}{\frac{\theta}{h}} = -\frac{\varsigma^2 - \theta p - \nu}{\theta} = -\frac{\epsilon}{\theta} + \frac{\nu}{\theta} = \mu - R_F + \frac{\nu}{\theta}$$

[35]For example, ρ denote the vector of covariance of risky asset's excess return with the quality of future investment opportunities (e.g., the risk free rate in Campbell and Viceira (1999) or risk premium in Campbell and Viceira (2001)). The investor's first order condition would then be:

$$\lambda\Sigma\boldsymbol{\omega}_i = \boldsymbol{\mu}_i - \boldsymbol{p} - \mathbf{1}R_F - \psi\rho.$$

actively update their portfolio every period (Gabaix, 2019). Generally speaking, this would result in our estimate of risk aversion being biased upwards because investors will appear as if they are insensitive to expected returns/fees. As such, investors would appear to be unwilling to take on additional risk after an increase in fund expected returns.[36] If we were to systematically over-estimate risk aversion, this would result in over-estimating investor optimism regarding fund returns because investors equate expected returns scaled by risk aversion to risk.

To help address the potential concern regarding inattentive investors and its potential impact on risk aversion, we separately examine the investment allocation decisions of participants in the year the 401(k) plan was first introduced. When a 401(k) plan is introduced, any allocation into non-target date funds by definition reflects an active choice of the participant. We discuss this robustness check in Section 5 and note that the estimated risk aversion appears roughly 20% lower in the year when the 401(k) plan was introduced. This suggests that some investors may be inattentive, but it does not appear to the driving factor of our estimate of risk aversion.

5 Estimates of Risk Preferences and Beliefs

Here we present our baseline estimates of risk aversion and beliefs and examine how they vary across investor demographics and characteristics. As documented in Section 3.1, we find substantial heterogeneity in investment portfolios across investors and that this heterogeneity is highly correlated with investor demographics. We use our model estimates to further under-

[36]Consider the a simple example where all investors in plan have the same beliefs at any given moment time, but only a fraction π of investors are attentive and update their portfolios. Also for convenience, assume that the factor loadings of the funds do not change over time such that $\Sigma = \Sigma_t \forall t$. Let $\boldsymbol{\omega}_t$ denote the optimal portfolio weights given that an investor updates her portfolio at time t:

$$\lambda \Sigma \boldsymbol{\omega}_t = \boldsymbol{\mu}_t - \boldsymbol{p_t} - \mathbf{1} R_{F,t}.$$

We do not observe the optimal portfolio weights at time t but rather some weighted function of the current and past optimal weights $\bar{\boldsymbol{\omega}}_t = \pi \boldsymbol{\omega}_t + \sum_{l=1}^{\infty} \pi (1-\pi)^l \boldsymbol{\omega}_{t-l}$. We can rewrite our estimation equation as:

$$\Sigma \bar{\boldsymbol{\omega}}_t = \frac{1}{\lambda} \left[\pi \left(\boldsymbol{\mu}_t - \boldsymbol{p}_t - \mathbf{1} R_{F,t} \right) + \sum_{l=1}^{\infty} \pi (1-\pi)^l \left(\boldsymbol{\mu}_{t-l} - \boldsymbol{p}_{t-l} - \mathbf{1} R_{F,t-l} \right) \right].$$

When we regress the term $\Sigma \bar{\omega}_t$ on the fund expense ratios $\boldsymbol{p}_t$ using two-stage least squares with our instrument $\boldsymbol{Z}_t$, our estimate $\hat{\frac{1}{\lambda}}$ will converge to

$$plim \, \hat{\frac{1}{\lambda}} = \frac{1}{\lambda} \left[\sum_{l=0}^{\infty} \pi (1-\pi)^l \frac{Cov(Z_t, p_{t-l})}{Cov(Z_t, p_t)} \right],$$

where we have assumed that our instrument Z_t is orthogonal to past changes in beliefs. If we further assume that $p_t = \phi p_{t-1} + \epsilon_t$ where $Cov(\epsilon_t, Z_t) = 0$ and $-1 < \phi < 1$, we can show that

$$plim \, \hat{\frac{1}{\lambda}} = \frac{1}{\lambda} \left[\sum_{l=0}^{\infty} \pi (1-\pi)^l \phi^{-l} \right] = \frac{1}{\lambda} \left(\frac{\pi}{(1 - \frac{1-\pi}{\phi})} \right) < \frac{1}{\lambda}.$$

stand why portfolios differ across investors and how much is driven by differences in investors' beliefs versus risk aversion.

5.1 Risk Preferences

We report our baseline model estimates corresponding to eq. (5) in Table 5. In specification (1), we keep the parameter θ_{mt} and consequently risk aversion fixed across 401(k) plans. In specifications (2)-(5), we allow θ_{mt} and risk aversion to vary across plans based on plan characteristics/demographics. In specifications (3) and (5), we also allow for arbitrary year-by-year variation in the mean level of θ by interacting fees with time dummy variables. For each specification, the left column reports the model estimates and standard errors. Recall that the parameter θ_{mt} corresponds to the negative inverse of risk aversion ($\theta_{mt} = -\frac{1}{\lambda_{mt}}$). For ease of interpretation, we report the corresponding estimates in terms of risk aversion and demographic interactions in the right column (λ).

We estimate mean risk aversion ranging from 3.6 to 5.2 across our main specifications. We also find that accounting for heterogeneity in risk aversion, as discussed further below, is important for explaining investment decisions. The interaction terms in Table 5 indicate how demographics are correlated with the parameter θ_{mt}. We find evidence that older plan participants behave as if they are more risk averse. The results in specification (2) of Table 5 indicate that a one standard deviation increase in age is associated with a 0.38 (8%) increase in risk aversion. Education is positively correlated with risk aversion. A one standard deviation increase in fraction with some college education is correlated with a 0.56 (12%) increase in risk aversion (specification 2, Table 5). Wealthier investors, as measured by median family income, tend to behave as if they are less risk averse, such that a one standard deviation increase in log income is correlated with a 0.47 (10%) decrease in risk aversion (specification 2, Table 5).

Lastly, in specifications (4) and (5) of Table 5 we allow risk aversion to vary in the year the 401(k) plan was first introduced. As discussed in Section 4.5, if investors are inattentive then they may appear more risk averse in the data than they actually are. Consistent with this, we find that investors behave as if their risk aversion is 0.89 (20%) lower in the year of inception (specification 4). Consequently, when constructing our estimates of risk aversion and beliefs in the remainder of our analysis, we set the variable $Existing\,401(k)\,Plan$ equal to zero to adjust for potential effects of inattention. In Appendix Table A7, we also show that we get similar estimates of beliefs and risk aversion if we restrict our sample to the first year each 401(k) plan was introduced.

Figure 6 displays the distribution of risk aversion over time where we allow the average level of risk aversion to vary over time (Table 5, specification 5). The solid red line displays the average risk aversion across plans and the dashed/dotted lines correspond to different quantiles of the distribution. The results suggest that risk aversion fell in 2010 as the economy was coming out of the global financial crisis and then peaked again in 2012 and 2013 around the time of the European sovereign debt crisis. Consistent with the estimates reported in Table 5, Figure 6 illustrates that there is substantial heterogeneity in risk aversion across plans/investors. Plans in the 90th percentile of the risk aversion distribution behave as if they are more than 25% more risk averse than plans in the 10th percentile of the risk aversion distribution. We find that this dispersion in risk aversion helps explain investors portfolio allocations in Section 5.3.

5.2 Investor Beliefs

Figure 7 displays the distribution of beliefs about the market return (δ_{mt}) over the period 2009-2019, where we allow risk aversion to vary across plans (corresponding to specification 5 in Table 5). The bright red solid line displays the average belief across plans over time. The results suggest that optimism remained relatively constant over the early part of our sample as the average investor expected the market return to be roughly 11%. Investors remained optimistic through 2017 and then the average expected return fell to roughly 7.4% in 2019. The average expected return over our sample is 9.6%, which is remarkably close to the realized excess return of the S&P 500 over this period. The compound annual growth rate (CAGR) of the excess return of the S&P 500 over the period 2009-2019 was roughly 10.7%.

There is also substantial heterogeneity in beliefs across plans. In Figure 7, we plot the 10th, 25th, 50th, 75th, and 90th percentile of expected returns in addition to the mean. Moving from the 10th to the 90th percentile of the distribution implies an increase in expected returns of roughly 5 percentage points in most years. For example, in 2011, the 10th percentile expected return is 8 pp and the 90th percentile is 14 pp. The standard deviation in expected market returns across plans within a year is 2.30 pp on average.

The differences in expected returns across plans are persistent. To demonstrate this, we calculate the average deviation from the within-year mean for each plan over time. Figure 8 displays the average plan-level deviation from the mean, i.e., the persistent cross-sectional variation in expected returns across plans. The standard deviation is 1.8pp, which is close to the plan-year standard deviation of 2.3. Thus, our estimates imply that relative pessimism and relative optimism about market returns are persistent features of retirement plans.

Note that the our analysis examines the cross-sectional dispersion in the average plan beliefs, where each plan is a collection of individuals. Given that median plan has more than 200 participants, and to the extent that there is variation in investor beliefs within plans, the dispersion shown in Figure 7 and Figure 8 could understate the individual-level dispersion in beliefs by an order of magnitude.

To better understand what drives heterogeneity in investor beliefs, we regress market beliefs (δ_{mt}) on a vector of plan characteristics (X_{mt}). Because risk aversion and beliefs tend to be positively correlated in the data ($R^2 = 0.20$; Figure 9) and risk aversion is a deterministic function of the covariates X_{mt}, we examine how the variation in market beliefs that is orthogonal to risk aversion (δ^*_{mt}) varies with plan characteristics. In other words, we examine how the covariates X_{mt} explain variation in beliefs that is orthogonal to risk aversion in the following regression:

$$\delta^*_{mt} = X'_{mt}\Gamma + \nu_{mt}. \tag{9}$$

Observations are at the plan-by-year level. The dependent variable δ^*_{mt} measures the residualized variation in expected market returns averaged across investors participating in plan m at time t that is orthogonal to risk aversion.[37] We control for the same set of industry and plan characteristics as in our previous analysis in Section 3.1.

Table 6 displays the estimates corresponding to eq. (9). We include year fixed effects in each specification. Columns (1)-(11) display univariate regressions and column (12) includes the full set of control variables. In general, we find that wealthier and more educated investors tend to have more optimistic expectations about the market. This helps explain why wealthier investors have higher equity participation rates. The results in column (2) indicate that a one standard deviation increase in ln(Income) is associated with a 0.17 pp increase in expected market returns. Similarly, a one standard deviation increase in the fraction of college educated individuals is associated with a 0.23 pp increase in expected market returns (Column 4, Table 6).

In contrast, we find that older investors, retirees, and minorities tend to have more pessimistic expectations about market returns. The results in column (12) indicate that a one standard deviation increase in the fraction of Hispanic (black) individuals is correlated with a 0.08 pp (0.10 pp) decrease in expected returns. These differences in market expectations could be driven by differences in trust (Guiso et al., 2008; Gennaioli et al., 2015) which may differ across ethnicities (Chiteji and Stafford, 2000).

We also find some evidence that participants' beliefs are shaped by their industry. The results in column (9) and (12) indicate that investors who work in riskier sectors, as measured by the equity beta of their sector, tend to have more optimistic beliefs. We look at this further by examining how beliefs about the market vary across sectors in Figure 10. The results suggest that there is substantial heterogeneity across sectors. At the median, investors from the most optimistic sector, Real Estate, expect the market return to be roughly 30 percent higher than investors from the least optimistic sector, Accommodation and Food Services (10.8% versus 8.3%). Investors in Real Estate also have meaninfully higher expected returns than those in Construction (10.8% versus 8.5%), despite having arguably similar risk profiles. It is interesting to examine how both beliefs and equity allocations vary across sectors by comparing Figures 3 and 10. To facilitate the comparison, panel (b) of Figure 10 sorts the sectors by median share allocated to U.S. equities. Though equity allocation and expected market returns are correlated, the correlation is far from perfect, suggesting the important role of variation in risk aversion across sectors as well. We also find evidence that there is substantial heterogeneity in beliefs

[37]We calculate δ^*_{mt} as the residual from the regression of δ_{mt} on the parameter θ_{mt},which corresponds to the negative inverse of risk aversion.

within a sector. The average interquartile range of beliefs within a sector is 3.1 pp. In other words, within a sector those investors in the 75th percentile of the beliefs distribution expect the market return to be roughly 40% higher than investors in the 25th percentile of the beliefs distribution.

An advantage in our setting is that we recover investors' beliefs about each investment option appearing in their plan, not just beliefs about the overall stock market. Figure 11 displays the estimated distributions of investors' expectations of returns across investors for each investment category. For every plan in every year, we compute category-level expected returns by averaging expected returns across all investment options available in each category, and we plot the distribution of category-level returns across plans and years. Investors' expectations of returns are the highest for small cap stock funds and are the lowest for bond funds. As with investor beliefs about the overall stock market (Figure 7), we find that expected returns are heterogeneous across investors for each investment category.

5.3 What Explains Holdings? Beliefs vs. Risk Aversion

Our results above indicate that there is substantial heterogeneity in beliefs and risk aversion across investors. We examine how dispersion in beliefs and risk aversion explain variation in equity exposure across plans in the following regression:

$$Equity\,Share_{mt} = \beta\lambda_{mt} + \gamma\delta_{mt} + \epsilon_{mt}. \tag{10}$$

Observations are at the plan-by-year level. The dependent variable $Equity\,Share_{mt}$ measures the share of assets in plan m that are invested in US equities. The dependent variables λ_{mt} and δ_{mt} measure the risk aversion and average market expectations of investors in plan m at time t.

Table 7 displays how dispersion in risk aversion and expectations explain 401(k) portfolios. The dependent variable in the regression specification displayed in columns (1) and (2) is the share of the portfolio held in equities (US and international equities), the dependent variable in columns (3) and (4) is the share held in US equities, and the dependent variable in column (5) and (6) is the share held in cash. To aid interpretation we also normalize risk aversion and investor beliefs such that each is mean zero and has a variance equal to one.

The results are intuitive and suggest that variation in beliefs and risk aversion both play important roles in explaining investor equity and cash allocations. The results in column (2) indicate that a one standard deviation increase in expected returns is correlated with a 13.7 pp (20% relative to the mean allocation) increase in an investor's equity allocation and a one standard deviation increase in risk aversion is correlated with a 7.4 pp (11% relative to the mean allocation) decrease in an investor's equity allocation. Conversely, an investor's expectations of the market return are negatively correlated with her cash holdings, and an investor's risk aversion is positively correlated with her cash holdings. The results also indicate that our simple two parameter model explains a fair amount of the variation in equity and portfolio holdings. Variation in beliefs and risk aversion explain 51% of the reduced-form variation in equity exposure.

Our framework also allows us to understand if the differences in equity exposure across investors are driven by differences in beliefs, risk aversion, or both. Thus, our findings provide a useful lens for understanding why portfolio allocations vary across investors, as documented in Section 3.1. For example, older investors have lower equity exposure because they are both more risk averse and more pessimistic. Individuals with more education allocate more towards equity because they have optimistic beliefs despite being more risk averse. Beliefs rather than risk aversion explain why equity allocation varies across ethnicities.

As an alternative way to illustrate the relative importance of heterogeneity in beliefs and risk aversion, we simulate allocations under counterfactual environments in which investors have identical beliefs, identical risk aversion, or both. For these counterfactuals, we use our method to calculate a single "average" expected return for each fund and an average risk aversion parameter, separately by year. We then calculate the optimal portfolio such that equation (3) is satisfied when replacing our estimated beliefs/risk aversion with the average values.

For the risk aversion parameter, we use the mean estimated value across plans, weighted by total plan assets. For expected returns, we aggregate fund balances across all plans and calculate the implied beliefs for each fund that would rationalize this aggregate portfolio under the average risk aversion parameter.[38] For the purposes of these counterfactuals, we only focus on plans with more than three investment options.

Figure 12 plots the densities of equity allocations across plans in 2016. The dashed line indicates the distribution of assets held in U.S. equity funds in our data. The solid line indicates the counterfactual distribution when removing heterogeneity in beliefs, i.e., assigning all investors identical fund-specific expected returns. The dotted line indicates the counterfactual distribution when assigning all investors identical values for risk aversion. Finally, the counterfactual distribution where we assign investors identical beliefs and risk aversion is given by the gray shaded area. To show the different counterfactuals on a more reasonable scale, the top of the density is visually cropped in the figure.

These counterfactual allocations indicate the importance of heterogeneity in beliefs and risk aversion in investor portfolio choice. Assigning investors identical beliefs greatly reduces the variation in equity allocations across plans. By comparison, assigning all investors the same risk aversion slightly increases the variation in equity allocations across plans. In this sample, the across-plan standard deviation in equity allocations is 0.132. With identical beliefs, the standard deviation falls to 0.072, but with average risk aversion, it increases to 0.141. Removing heterogeneity in beliefs and risk aversion together further reduces variation across plans,

[38] Alternatively we could calculate beliefs using the average estimated belief (across investors) for each asset using our estimates from Section 5.2. The correlation between this measure of implied beliefs and the average estimated belief (across investors) in 2016 is 0.91.

lowering the standard deviation in equity allocations to 0.043. The residual variation when investors have identical beliefs and risk aversion is due to differences in menus across plans. Our estimates indicate that both heterogeneity in beliefs and risk aversion are important; however, these simulations suggest that variation in beliefs plays a bigger role in driving variation across plans.

5.4 Alternative Specifications and Robustness

We consider several alternative specifications to assess the robustness of the estimated parameters. First, we re-estimate the model to include target date funds, which are excluded from our baseline analysis. Second, to account for potential inertia in investor behavior, we estimate the model using only new plans. For all of these specifications, we find very similar estimates of risk aversion and expected returns. Results are reported in Table A7. The mean risk aversion we estimate in these alternative models is nearly identical to our baseline estimate (3.55 and 3.56 vs. 3.55). The mean expected return ranges from 9.7 to 9.9, similar to our baseline estimate of 9.6. As shown in panel (b), individual estimates of expectations and risk aversion are positively and significantly correlated with the baseline specification.

In addition, we consider alternative measures of risk based on both simplified and more extensive factor structures, which we describe in Section 4.3.[39] As in the above specifications, we find that the estimates of risk aversion and beliefs are highly correlated with our baseline estimates.

6 Evidence on the Formation of Beliefs

Investor beliefs play a critical role in determining investor portfolios and vary substantially across investors. Here, we provide insight into how beliefs are formed across investors.

A large previous literature documents that investors extrapolate beliefs from past returns and experiences. Our unique setting provides additional insight into how investors extrapolate their beliefs on two dimensions. First, we find that investors extrapolate their beliefs from past fund returns. Given that past returns are often a salient feature of 401(k) brochures/documents, investors are likely to extrapolate by looking at last year's returns. Using variation in 401(k) menus over time, we also show that investors extrapolate from past returns for both existing funds and funds newly added to their menus. Thus, the extrapolation we document cannot be explained by a lack of rebalancing.

[39] We construct the factors for our simplified measure of risk by forming equal weighted portfolios based on the broad BrightScope categories reported in Table 1a. While the baseline and simplified measures of risk are highly correlated (0.94), the standard deviation of our simplified measure of risk is roughly 40% smaller than the standard deviation of our baseline measure of risk. This helps explain why we estimate higher average risk aversion (7.63) with our simplified measure of risk relative to our baseline estimates.

Second, to understand the role of experience in shaping extrapolation, we examine how investors extrapolate their beliefs based on local conditions *above and beyond* what is available in terms of aggregate information. Consider an investor i at period t who believes that the future state of the stock market, z_{t+1}, is predictable. The investor forms expectations based on aggregate variables, $\boldsymbol{w}_t$, which are common knowledge and, as above, may include past market returns. The investor also takes into account idiosyncratically observed variables, $\boldsymbol{\iota}_{it}$, forming the forecast

$$z_{i(t+1)} = \Theta \boldsymbol{w}_t + \Xi \boldsymbol{\iota}_{it}, \tag{11}$$

which yields an expected return $\frac{z_{i(t+1)} - z_t}{z_t}$. The heterogeneity in beliefs we document above suggests that there is substantial variation in $z_{i(t+1)}$ across investors. To provide evidence of the role of idiosyncratic experiences ($\boldsymbol{\iota}_t$) in forming beliefs, we examine how market expected returns reflect local economic conditions and past performance of an investor's employer, while controlling for aggregate information ($\boldsymbol{w}_t$), which can be captured with time fixed effects. We find that local economic conditions and employer past performance are positively correlated with beliefs about market returns, suggesting that investors form broader beliefs about market returns based on individualized experiences.

Lastly, we assess the rationality of investor beliefs. The above evidence, which documents systematic and predictable drivers of heterogeneity of beliefs, suggests that a standard rational expectations model may not capture the investment behavior across households. Investor beliefs are correlated with observable characteristics such as wealth and income, and appear to depend on past market returns as well as on recent employer performance. We find evidence consistent with the vast prior literature suggesting that investor forecasts violate full information rational expectations. Forecast errors are predictable and forecast revisions, measured by changes in investor expectations, are correlated with future forecast errors.

6.1 Extrapolation from Fund Returns

We examine how investors form their beliefs for a particular fund based on the fund's return over the previous year. We estimate the regression:

$$\bar{\mu}_{kt}^{(m)} = \rho Ret_{kt-1} + \upsilon_{kt}. \tag{12}$$

Observations are at the investment option-by-plan-by-year level. The dependent variable measures the average participant in plan m's expected return of fund k ($\bar{\mu}_{kt}^{(m)}$). The independent variable Ret_{kt+1} measures the past monthly return of investment option k averaged over year $t-1$ to t and is annualized. Table 8 displays the estimates corresponding to eq. (12). We examine extrapolation across three different subsets of the data: (i) the full data set in columns (1), and (4); (ii) fund-by-plan observations in the first year the fund was added to the plan in column (2);[40] and (iii) fund-by-plan observations corresponding to the first year a 401(k) plan was introduced in column (3). Samples (ii) and (iii) allow us to examine how investors extrapolate their beliefs about funds they have not previously held in their 401(k).

We find evidence that investors extrapolate their beliefs from past returns. The results in columns (2)-(4) indicate that investors extrapolate their beliefs from past returns for funds they did not hold in the past. The results in column (2) indicate that a ten percentage point increase in last year's return is correlated with an 0.16 pp increase in expected returns. In column (4) we interact past returns with the dummy variable $New\,Investment_{kt}$, which indicates whether the fund was added to the 401(k) menu in year t. We find a small statistically insignificant coefficient which indicates that investors extrapolate in the same way for both new and existing funds. The results in columns (2)-(4) show that the extrapolation we observe is not simply a function of investor inattention or inertia in portfolio rebalancing. In Appendix Table A10, we show that we find similar results if examine portfolio weights rather than beliefs.

6.2 Extrapolation from Local Economic Conditions

Investors may also extrapolate from local economic conditions. We examine the relationship between investors' beliefs and local economic conditions in the following regression:

$$\delta_{mt} = Local\,Economic\,Conditions'_{mt}\Gamma + \mu_t + \mu_m + \varepsilon_{mt}. \tag{13}$$

Observations are at the plan-by-year level. The dependent variable δ_{mt} measures the expected market return averaged across investors in plan m at time t. The term $Local\,Economic\,Conditions_{mt}$ is a vector of county-by-year level measures of economic conditions including: GDP growth, business establishment growth, annual home price growth, and population growth.[41] We also control for year (μ_t) and plan (μ_m) fixed effects. Thus we measure how, conditional on aggregate macroeconomic conditions, changes in local economic condition are correlated with changes in macroeconomic beliefs

We report the estimates corresponding to eq. (13) in Table 9. In each specification, we find a strong positive relationship between local macroeconomic conditions and investors' beliefs about stock market returns. The results in column (1) indicate that a 1% increase in county population is correlated with a 0.13pp increase in expected returns. Similarly, the results in column (3) indicate that a 10% increase county home prices is associated with a 0.22pp increase in expected returns. We find a positive relationship between each measure of local economic activity and market expectations, even when we use within plan variation. In the Appendix, we show that these effects spillover to equity holdings as well. Overall, these results suggest that

[40] To keep the sample distinct from sample (iii), we exclude all fund-by-year observations when the 401(k) plan is introduced.

[41] We measure home price growth using data from the FHFA, GDP growth from the Bureau of Economic Analysis, establishment growth from the County Business Patterns, and population growth from the Census.

idiosyncratic experiences may drive differences in expected returns across investors, potentially through how they shape forecasts of future returns (eq. (11)).

6.3 Extrapolation from Employer Performance

An advantage in our setting is that we observe details on the investor's employer, the fund sponsor. This allows us to explore how investors' beliefs depend on their employment. Using the sponsor's EIN number, we link our BrightScope 401(k) data with balance sheet, income statement, and market return data from Compustat and CRSP.

We examine the relationship between the financial performance of an investor's employer and the investor's beliefs in the following regression:

$$\delta_{mt} = \varphi Performance_{mt} + \mu_t + \mu_m + \eta_{mt}. \tag{14}$$

Observations are at the plan-by-year level where we restrict the data set to those plans where the sponsor is publicly traded. The dependent variable δ_{mt} measures the expected market return averaged across investors in plan m at time t. The independent variable $Performance_{mt}$ measures the financial performance of plan sponsor m at time t. We measure firm performance in terms of last year's annual stock market return, sales growth, investment, and employment growth.

We report the estimates corresponding to eq. (14) in Table 10. Consistent with our previous results, we find that beliefs are highly correlated with local conditions. In each specification we document a positive and significant relationship between sponsor performance and participants' expectations about the market. The results are robust to the inclusion of both plan fixed effects (odd columns) and industry-by-year fixed effects (even columns). Including industry-by-year fixed effects allows us to effectively compare the beliefs of two investors working in the same industry at the same time but for different firms. In columns (1) and (2) we find that investors become more optimistic about the market following strong performance of their employer. The effect is marginally stronger when we include industry-by-year fixed effects which suggests that investors are more sensitive to industry or risk adjusted return than absolute returns. The results in column (4) indicate that investors become 0.18 pp more optimistic about the expected return of the market following a 10% increase in investment. Similarly, we estimate that a one standard deviation increase in sales growth (24%) is associated with a 0.10 pp increase in expected return of the market (column 6). In the Appendix, we document that we find a similar positive relationship between equity holdings and employer performance. Overall, this suggests that investors may misattribute the performance of their employer to the performance of the economy more generally, or they use this more local experience to form an idiosyncratic forecast of future market returns, as in eq. (11).

6.4 Are Beliefs Rational?

Lastly, we examine the rationality of investor beliefs by examining forecast errors. The previous results already provide suggestive evidence that investor beliefs are irrational. The unpredictability of forecast errors is a necessary condition for rational forecasts. We construct forecast errors at the plan-by-investment-by-year level as:

$$\varepsilon_{mkt+1} = Ret._{kt+1} - \bar{\mu}_{kt}^{(m)} \tag{15}$$

where $Ret._{kt+1}$ measures the monthly return of investment option k averaged over year t to $t+1$ and is annualized. The term $\bar{\mu}_{kt}^{(m)}$ is the average participant in plan m's expected return of fund k. We test the predictability of forecast errors in the following regression model:

$$\varepsilon_{mkt+1} = \alpha_0 + \alpha_1 X_{mkt} + \eta_{mkt+1}. \tag{16}$$

Observations are at the investment option-by-plan-by-year level. The vector X_{mkt} consists of a number of investment option and plan characteristics. We examine how forecast errors vary with past forecast errors, past fund returns, and changes in investor expectations.

Table 11 displays the estimation results corresponding to eq. (16). In short, we find overwhelming evidence that forecast errors are predictable. The results in column (1) indicates that forecast errors are persistent. We also find that investors tend to over predict fund returns following past positive fund returns (columns 3 and 4). This is consistent with our finding, discussed in the proceeding section, that investors extrapolate from previous returns. We also find that changes in beliefs are negatively correlated with future forecast errors. This test is in a similar vein as the test developed in (Coibion and Gorodnichenko, 2015) and employed in (Bordalo et al., 2018) where the researchers examine how forecast errors correlate with forecast revisions. The negative relationship between changes in beliefs and future forecast errors suggests that investors overreact to news.

One might expect that these patterns are driven by inexperience in financial markets. In the Appendix, we replicate our analysis where we restrict our analysis to those plan sponsors in the finance and insurance sector (NAICS 52). Similar to our baseline results, we find that the beliefs of investors working in the financial sector are extrapolative, violate full information rational expectations, and tend to overreact to news.

7 Conclusion

We examine how households allocate their 401(k) portfolios. Allocations vary dramatically across plans and vary in systematic ways with participant and employer characteristics. For example, plans with more educated participants tend to hold more of their portfolio in equities and less in cash. In contrast, the investment options available to plan participants do not vary systematically with participant characteristics.

To understand the patterns we document, we propose a framework for estimating investor beliefs and risk aversion. By measuring how investors re-optimize their portfolios in response to exogenous changes in investment fees, we are able to separately identify risk aversion from beliefs. Studying 401(k) plan allocations, where investors choose from a preset menu of investment options with variations in expense ratios, offers an ideal setting for our framework.

We find that there is substantial heterogeneity in both risk aversion and beliefs across investors. The differences in expectations and risk aversion are correlated with observable investor characteristics and help explain the heterogeneity in asset allocation across plans. For example, our results suggest that differences in beliefs, rather than risk aversion, help explain why educated investors tend to hold more equities and less cash. Counterfactual simulations suggest that heterogeneity in beliefs drives the majority of variation in equity allocations.

An important feature of our model is that we do not impose any restrictions on the rationality of beliefs. In fact, we find that investor beliefs violate rational expectations and tend to overreact to recent news. Investors extrapolate their beliefs from both past fund returns and from individualized experience based on local economic conditions and employer performance.

Our results also highlight the importance of accounting for and understanding heterogeneity in both beliefs and risk aversion. We show that both sources of heterogeneity play important roles in explaining equity participation rates across investors and potentially have important implications for asset prices. Our framework can also be easily applied in other settings to provide insight about investor beliefs and risk aversion, which could be particularly valuable when survey data is unavailable.

References

Ameriks, J. and S. P. Zeldes (2004). How do household portfolio shares vary with age. *Working paper*.

Amromin, G. and S. A. Sharpe (2014). From the horse's mouth: Economic conditions and investor expectations of risk and return. *Management Science 60*(4), 845–866.

Bacchetta, P., E. Mertens, and E. Van Wincoop (2009). Predictability in financial markets: What do survey expectations tell us? *Journal of International Money and Finance 28*(3), 406–426.

Bach, L., Laurent E. Calvet, and Paolo Sodini (2020). Rich pickings? Risk, return, and skill in household wealth. *American Economic Review 110*(9), 46.

Badarinza, C., J. Y. Campbell, and T. Ramadorai (2016). International comparative household finance. *Annual Review of Economics 8*, 111–144.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review 103*(6), 2499–2529.

Bekaert, G., K. Hoyem, W.-Y. Hu, and E. Ravina (2017). Who is internationally diversified? evidence from the 401 (k) plans of 296 firms. *Journal of Financial Economics 124*(1), 86–112.

Ben-David, I., J. R. Graham, and C. R. Harvey (2013). Managerial miscalibration. *The Quarterly Journal of Economics 128*(4), 1547–1584.

Benartzi, S. (2001). Excessive extrapolation and the allocation of 401 (k) accounts to company stock. *The Journal of Finance 56*(5), 1747–1764.

Benartzi, S. and R. Thaler (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives 21*(3), 81–104.

Benartzi, S. and R. H. Thaler (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review 91*(1), 79–98.

Benetton, M. (2021). Leverage regulation and market structure: A structural model of the UK mortgage market. *The Journal of Finance 76*(6), 2997–3053.

Benetton, M. and G. Compiani (2021). Investors beliefs and cryptocurrency prices. *Working paper*.

Berk, J. B. and J. H. van Binsbergen (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics 119*, 1–23.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2009). *5. The Importance of Default Options for Retirement Saving Outcomes: Evidence from the United States*. University of Chicago Press.

Bhattacharya, V. and G. Illanes (2021). The design of defined contribution plans. *Working paper*.

Black, S. E., P. J. Devereux, P. Lundborg, and K. Majlesi (2018). Learning to take risks? the effect of education on risk-taking in financial markets. *Review of Finance 22*(3), 951–975.

Bodie, Z., R. C. Merton, and W. F. Samuelson (1992). Labor supply flexibility and portfolio choice in a life cycle model. *Journal of economic dynamics and control 16*(3-4), 427–449.

Bordalo, P., G. Burro, K. Coffman, N. Gennaioli, and A. Shleifer (2022). Imagining the future: Memory, simulation and beliefs about covid. *Working Paper*.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2018). Over-reaction in macroeconomic expectations. *Working paper*.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2019). Diagnostic expectations and stock returns. *The Journal of Finance 74*(6), 2839–2874.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *The Journal of Finance 73*(1), 199–227.

Bretscher, L., L. Schmid, I. Sen, and V. Sharma (2020). Institutional corporate bond pricing. *Working paper*.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018). Beyond the balance sheet model of banking: Implications for bank regulation and monetary policy. *Working paper*.

Calvet, L. E., J. Y. Campbell, F. J. Gomes, and P. Sodini (2019). The cross-section of household preferences. Working paper.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy 115*(5), 707–747.

Campbell, J. Y. (2006). Household finance. *The journal of finance 61*(4), 1553–1604.

Campbell, J. Y. and L. Viceira (2001). Who should buy long-term bonds? *American Economic Review 91*(1), 99–127.

Campbell, J. Y. and L. M. Viceira (1999). Consumption and portfolio decisions when expected returns are time varying. *The Quarterly Journal of Economics 114*(2), 433–495.

Campbell, J. Y., L. M. Viceira, L. M. Viceira, et al. (2002). *Strategic asset allocation: portfolio choice for long-term investors*. Clarendon Lectures in Economic.

Carroll, G. D., J. J. Choi, D. Laibson, B. C. Madrian, and A. Metrick (2009). Optimal defaults and active decisions. *The quarterly journal of economics 124*(4), 1639–1674.

Case, K. E., R. J. Shiller, and A. K. Thompson (2012). What have they been thinking? Homebuyer behavior in hot and cold markets. *Brookings Papers on Economic Activity*, 265.

Chen, X., L. P. Hansen, and P. G. Hansen (2020). Robust identification of investor beliefs. *Working paper*.

Chiteji, N. S. and F. P. Stafford (2000). Asset ownership across generations. *Working paper*.

Choi, J. J. (2015). Contributions to defined contribution pension plans. *Annual Review of Financial Economics 7*, 161–178.

Choi, J. J., D. Laibson, B. C. Madrian, and A. Metrick (2002). Defined contribution pensions: Plan rules, participant choices, and the path of least resistance. *Tax Policy and the Economy 16*, 67–113.

Choi, J. J., D. Laibson, B. C. Madrian, A. Metrick, and J. M. Poterba (2007). *2. For Better or for Worse: Default Effects and 401(k) Savings Behavior*, pp. 81–126. University of Chicago Press.

Choi, J. J. and A. Z. Robertson (2020). What matters to individual investors? evidence from the horse's mouth. *The Journal of Finance 75*(4), 1965–2020.

Choukhmane, T. and T. de Sliva (2022). What drives investors portfolio choices? separating risk preferences from frictions. *Working paper*.

Cocco, J. F. (2005). Portfolio choice in the presence of housing. *The Review of Financial Studies 18*(2), 535–567.

Cocco, J. F., F. J. Gomes, and P. J. Maenhout (2005). Consumption and portfolio choice over the life cycle. *The Review of Financial Studies 18*(2), 491–533.

Cochrane, J. H. (2022). Portfolios for long-term investors. *Review of Finance 26*(1), 1–42.

Coibion, O. and Y. Gorodnichenko (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy 120*(1), 116–159.

Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review 105*(8), 2644–78.

d'Arienzo, D. (2020). Maturity increasing overreaction and bond market puzzles. *Working paper*.

Das, S., C. M. Kuhnen, and S. Nagel (2020). Socioeconomic status and macroeconomic expectations. *The Review of Financial Studies 33*(1), 395–432.

Di Maggio, M., M. Egan, and F. Franzoni (2021). The value of intermediation in the stock market. *Journal of Financial Economics*.

Dick, A. A. (2008). Demand estimation and consumer welfare in the banking industry. *Journal of Banking & Finance 32*(8), 1661–1676.

Du, W., S. Gadgil, M. B. Gordy, and C. Vega (2019). Counterparty risk and counterparty choice in the credit default swap market. *Working paper*.

Duflo, E., W. Gale, J. Liebman, P. Orszag, and E. Saez (2006). Saving incentives for low-and middle-income families: Evidence from a field experiment with H&R Block. *The Quarterly Journal of Economics 121*(4), 1311–1346.

Dworak-Fisher, K. (2011). Matching matters in 401 (k) plan participation. *Industrial Relations: A Journal of Economy and Society 50*(4), 713–737.

Egan, M. (2019). Brokers vs. retail investors: Conflicting interests and dominated products. *The Journal of Finance 74*(3), 1217–1260.

Egan, M., A. Hortaçsu, and G. Matvos (2017). Deposit competition and financial fragility: Evidence from the us banking sector. *American Economic Review 107*(1), 169–216.

Egan, M., S. Lewellen, and A. Sunderam (2022). The cross-section of bank value. *The Review of Financial Studies 35*(5), 2101–2143.

Egan, M. L., A. MacKay, and H. Yang (Forthcoming). Recovering investor expectations from demand for index funds. *Review of Economic Studies*.

Fagereng, A., C. Gottlieb, and L. Guiso (2017). Asset market participation and portfolio choice over the life-cycle. *The Journal of Finance 72*(2), 705–750.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020). Heterogeneity and persistence in returns to wealth. *Econometrica 88*(1), 115–170.

Farhi, E. and S. Panageas (2007). Saving and investing for early retirement: A theoretical analysis. *Journal of Financial Economics 83*(1), 87–121.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 261–343. Elsevier.

Gennaioli, N., Y. Ma, and A. Shleifer (2016). Expectations and investment. *NBER Macroeconomics Annual 30*(1), 379–431.

Gennaioli, N. and A. Shleifer (2018). *A crisis of beliefs: Investor psychology and financial fragility*. Princeton University Press.

Gennaioli, N., A. Shleifer, and R. Vishny (2015). Money doctors. *The Journal of Finance 70*(1), 91–114.

Ghosh, A., A. G. Korteweg, and Q. Xu (2020). Recovering heterogeneous beliefs and preferences from asset prices. *Working Paper*.

Ghosh, A. and G. Roussellet (2020). Identifying beliefs from asset prices. *Working paper*.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus (2021, May). Five facts about beliefs and portfolios. *American Economic Review 111*(5), 1481–1522.

Gomes, F., M. Haliassos, and T. Ramadorai (2020). Household Finance. *Journal of Economic Literature, forthcoming*.

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

Guiso, L., T. Jappelli, and D. Terlizzese (1996). Income risk, borrowing constraints, and portfolio choice. *The American Economic Review*, 158–172.

Guiso, L., P. Sapienza, and L. Zingales (2008). Trusting the stock market. *the Journal of Finance 63*(6), 2557–2600.

Haddad, V., P. Huebner, and E. Loualiche (2021). How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing. *Working paper*.

Hastings, J., A. Hortaçsu, and C. Syverson (2017). Sales force and competition in financial product markets: the case of mexico's social security privatization. *Econometrica 85*(6), 1723–1761.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In *The Economics of New Goods*, pp. 207–248. University of Chicago Press.

Heaton, J. and D. Lucas (2000). Portfolio choice and asset prices: The importance of entrepreneurial risk. *The Journal of Finance 55*(3), 1163–1198.

Heipertz, J., A. Ouazad, and R. Rancière (2019). The transmission of shocks in endogenous financial networks: A structural approach. *Working paper*.

Hortaçsu, A. and C. Syverson (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *The Quarterly Journal of Economics 119*(2), 403–456.

Huberman, G. and W. Jiang (2006). Offering versus choice in 401 (k) plans: Equity exposure and number of funds. *The Journal of Finance 61*(2), 763–801.

Jensen, C. S., D. Lando, and L. H. Pedersen (2019). Generalized recovery. *Journal of Financial Economics 133*(1), 154–174.

Koijen, R. S., R. J. Richmond, and M. Yogo (2019). Which investors matter for global equity valuations and expected returns? *Working paper*.

Koijen, R. S. and M. Yogo (2016). Shadow insurance. *Econometrica 84*(3), 1265–1287.

Koijen, R. S. and M. Yogo (2019a). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

Koijen, R. S. and M. Yogo (2019b). Exchange rates and asset prices in a global demand system. *Working paper*.

Koijen, R. S. and M. Yogo (2022). The fragility of market risk insurance. *The Journal of Finance 77*(2), 815–862.

Kronlund, M., V. K. Pool, C. Sialm, and I. Stefanescu (2021). Out of sight no more? the effect of fee disclosures on 401 (k) investment allocations. *Journal of Financial Economics 141*(2), 644–668.

Kuhnen, C. M. and A. C. Miu (2017). Socioeconomic status and learning from financial information. *Journal of Financial Economics 124*(2), 349–372.

Lynch, A. W. and S. Tan (2011). Labor income dynamics at business-cycle frequencies: Implications for portfolio choice. *Journal of Financial Economics 101*(2), 333–359.

Madrian, B. C. and D. F. Shea (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly journal of economics 116*(4), 1149–1187.

Malmendier, U. (2021). Experience effects in finance: Foundations, applications, and future directions. *Review of Finance 25*(5), 1339–1363.

Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics 126*(1), 373–416.

Malmendier, U. and S. Nagel (2015). Learning from inflation experiences. *The Quarterly Journal of Economics 131*(1), 53–87.

Malmendier, U., D. Pouzo, and V. Vanasco (2020). Investor experiences and financial market dynamics. *Journal of Financial Economics 136*(3), 597–622.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

Martin, I. and S. Ross (2019). Notes on the yield curve. *Journal of Financial Economics 134*(3), 689–702.

Meeuwis, M. (2019). Wealth fluctuations and risk preferences: Evidence from U.S. investor portfolios. *Working Paper*.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, 247–257.

Nagel, S. and Z. Xu (2019). Asset pricing with fading memory. *Working paper*.

Parker, J. A., A. Schoar, and Y. Sun (2020). Retail financial innovation and stock market dynamics: The case of target date funds. *Working paper*.

Pool, V. K., C. Sialm, and I. Stefanescu (2016). It pays to set the menu: Mutual fund investment options in 401 (k) plans. *The Journal of Finance 71*(4), 1779–1812.

Pool, V. K., C. Sialm, and I. Stefanescu (2020). Mutual fund revenue sharing in 401 (k) plans. *Working paper*.

Robles-Garcia, C. (2019). Competition and incentives in mortgage markets: The role of brokers. *Working paper*.

Ross, S. (2015). The recovery theorem. *The Journal of Finance 70*(2), 615–648.

Shumway, T., M. Szefler, and K. Yuan (2009). The information content of revealed beliefs in portfolio holdings. *Working paper*.

Sialm, C., L. T. Starks, and H. Zhang (2015). Defined contribution pension plans: Sticky or discerning money? *The Journal of Finance 70*(2), 805–838.

Sinai, T. and N. S. Souleles (2005). Owner-occupied housing as a hedge against rent risk. *The Quarterly Journal of Economics 120*(2), 763–789.

Storesletten, K., C. I. Telmer, and A. Yaron (2004). Cyclical dynamics in idiosyncratic labor market risk. *Journal of Political Economy 112*(3), 695–717.

Van Rooij, M., A. Lusardi, and R. Alessie (2011). Financial literacy and stock market participation. *Journal of Financial Economics 101*(2), 449–472.

Viceira, L. M. (2001). Optimal portfolio choice for long-horizon investors with nontradable labor income. *The Journal of Finance 56*(2), 433–470.

Vissing-Jorgensen, A. (2003). Perspectives on behavioral finance: Does "irrationality" disappear with wealth? evidence from expectations and actions. *NBER Macroeconomics Annual 18*, 139–194.

Wachter, J. A. and M. Yogo (2010). Why do household portfolio shares rise in wealth? *The Review of Financial Studies 23*(11), 3929–3965.

Wang, Y., T. M. Whited, Y. Wu, and K. Xiao (2018). Bank market power and monetary policy transmission: Evidence from a structural estimation. *Working paper*.

Xiao, K. (2020). Monetary transmission through shadow banks. *The Review of Financial Studies 33*(6), 2379–2420.

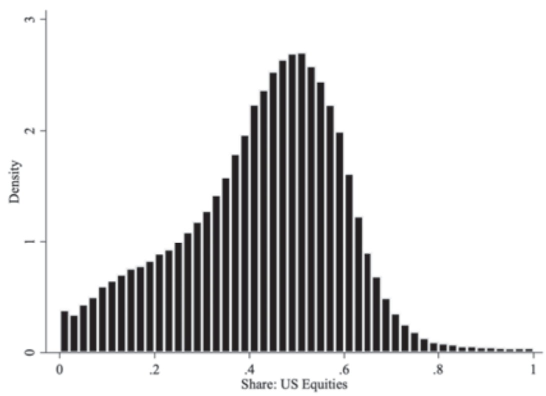
Yao, R. and H. H. Zhang (2005). Optimal consumption and portfolio choices with risky housing and borrowing constraints. *The Review of Financial Studies 18*(1), 197–239.

Yogo, M., A. Whitten, and N. Cox (2021). Financial inclusion across the united states. *Working paper*.

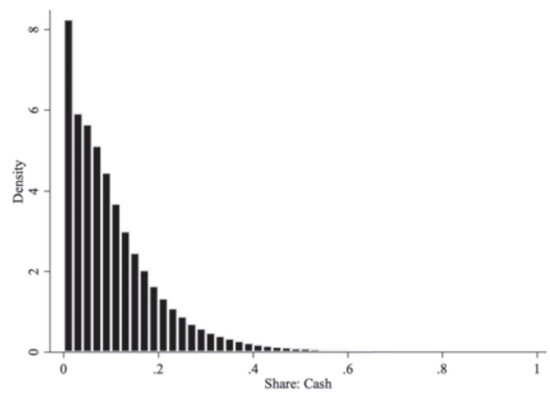
Tables and Figures

Figure 1: Distribution of Holdings

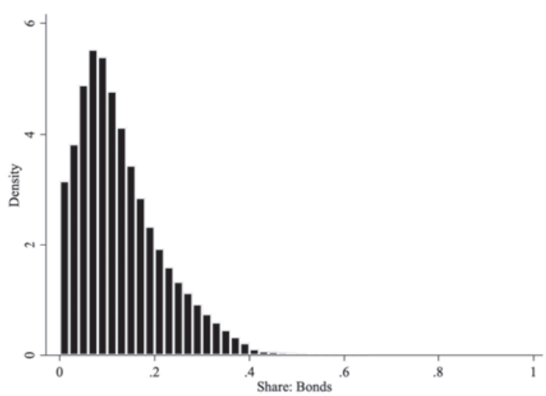
(a) Equities



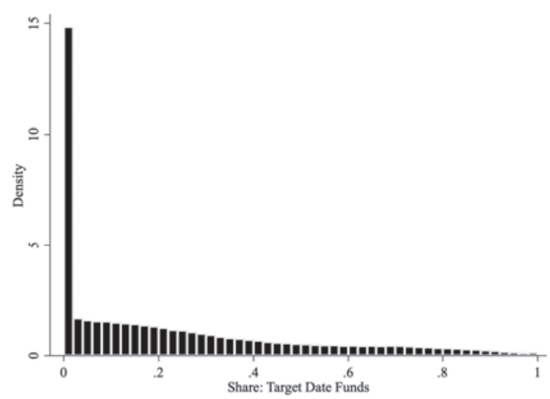
(b) Cash



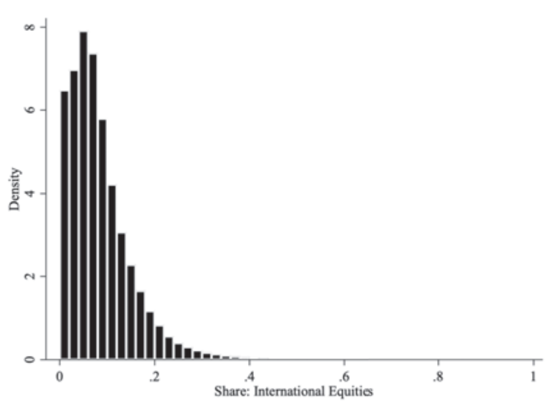
(c) Bonds



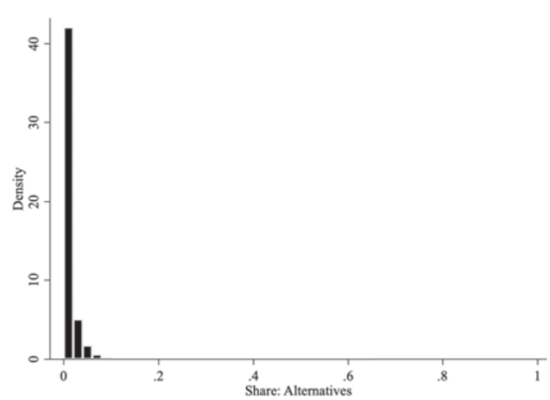
(d) Target-Date Funds



(e) International Assets



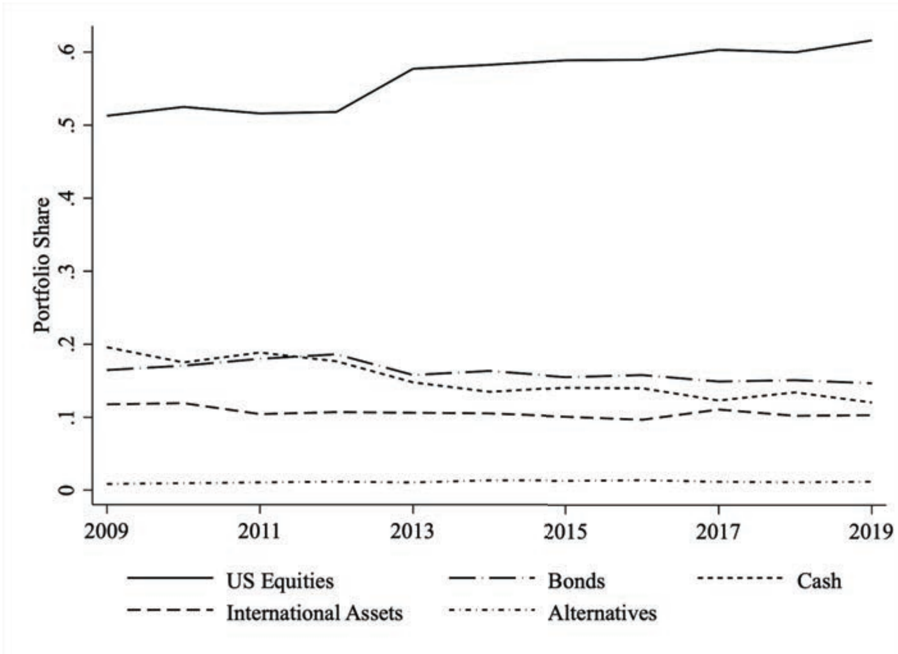
(f) Alternatives



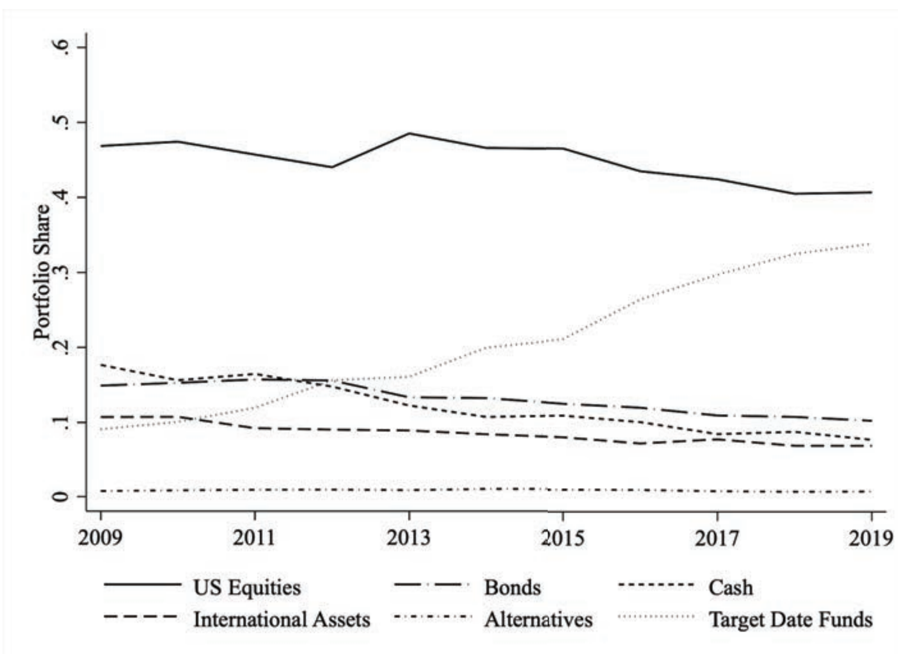
Notes: Figure 1 displays the distribution of holdings across 401(k) plans. Observations are at the plan-by-year level over the period 2009-2019 for those plans with at least five investment options.

Figure 2: Holdings Over Time

(a) Holdings Over Time, Excluding Target Date Funds

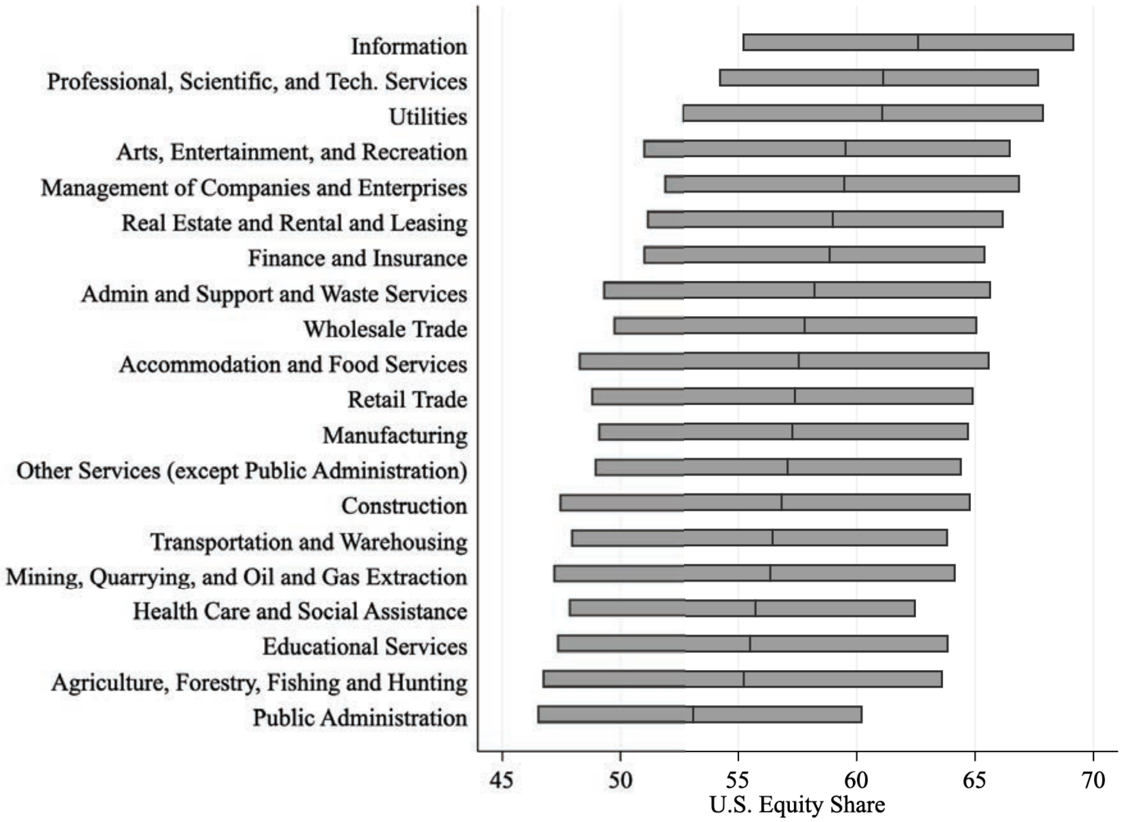


(b) Holdings Over Time, Including Target Date Funds



Notes: Figure 2 displays the equal-weighted average holdings across plans over the period 2009-2019. In panel (a) we calculate portfolio shares excluding target date funds. In panel (b) we calculate portfolio shares including target date funds.

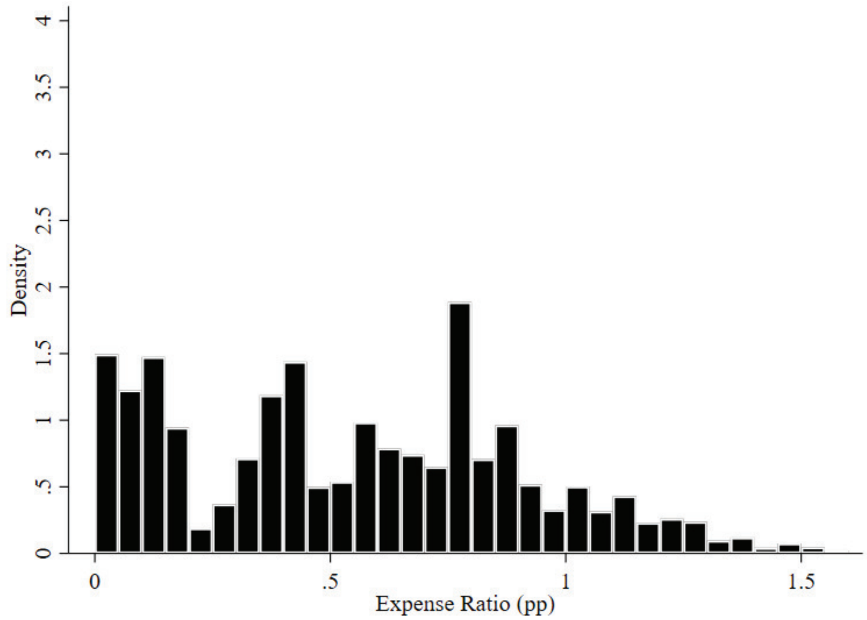
Figure 3: Equity Allocation by Sector of Employment



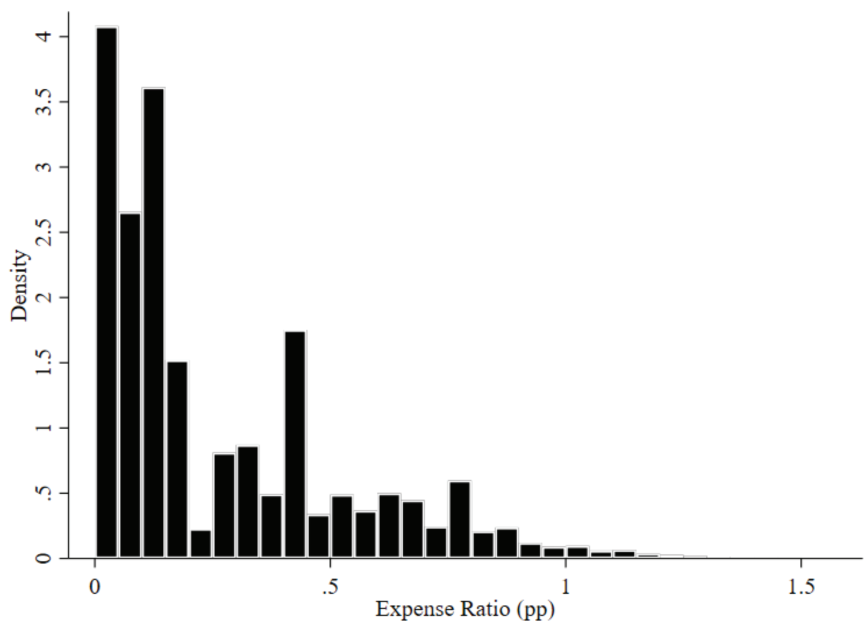
Notes: Figure 3 displays the distribution of US equity allocations (i.e., share of plan assets held in US equities) across sectors (2-digit NAICS). The horizontal gray bars cover the 25th to 75th percentiles, and the short vertical lines indicate medians. When computing the share held in US equities, we drop all target date fund assets and assume that remaining non-target date allocation funds hold 60% in US equities. Observations are at the plan-by-year level over the period 2009-2019.

Figure 4: Fund Expenses

(a) Fund Expenses (Equal Weighted)



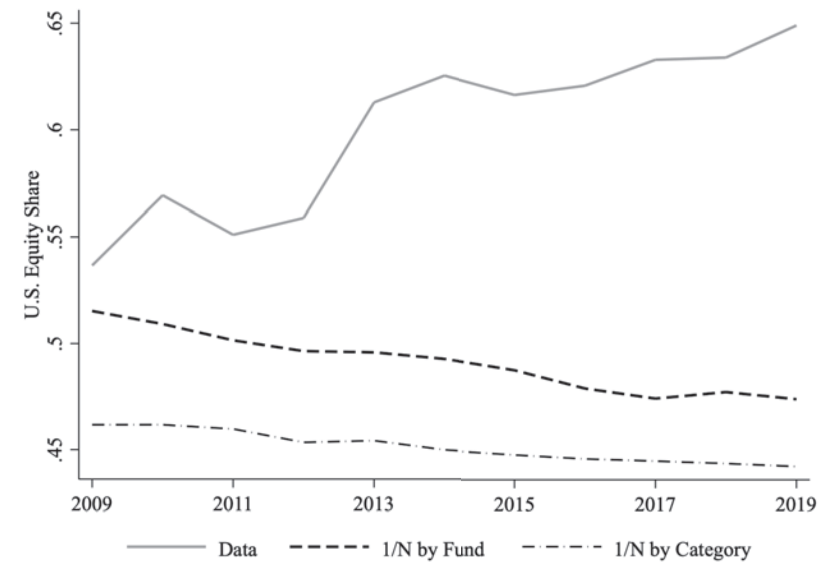
(b) Fund Expenses (AUM Weighted)



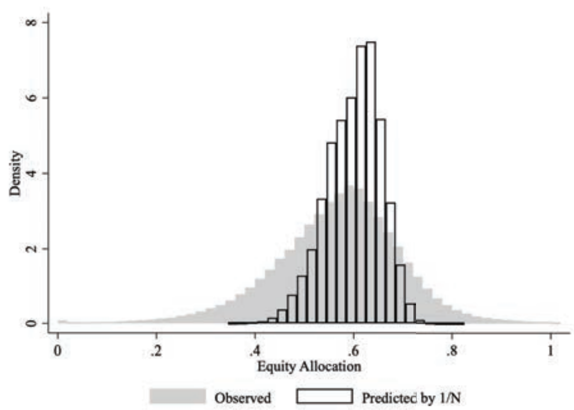
Notes: Figure 4 displays the distribution of fund expenses. Observations are at the fund-by-plan level as of 2019 as reported by BrightScope. Panel (a) displays the equal weighted distribution of fund expenses. Panel (b) displays the asset weighted distribution of fund expenses.

Figure 5: Equity Allocations and Naive Diversification

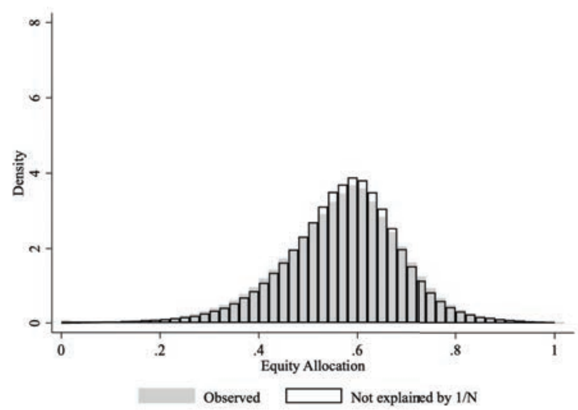
(a) Observed Allocations Compared to Naive Diversification



(b) Observed vs Predicted Allocations

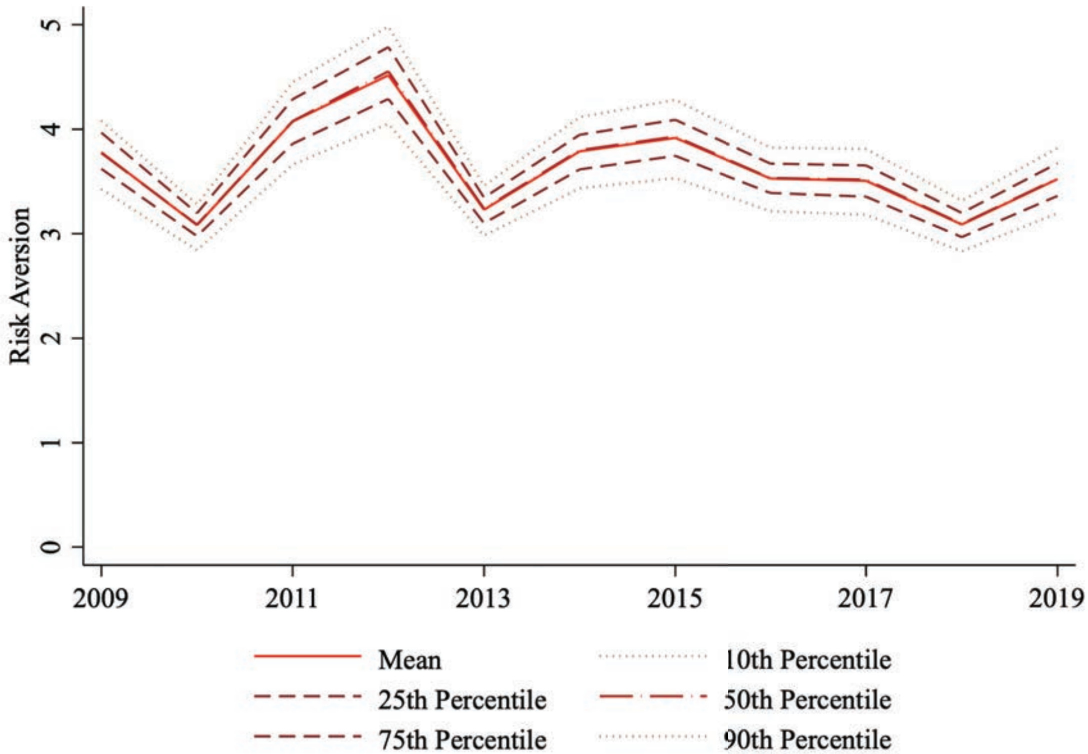


(c) Observed vs Residualized Allocations



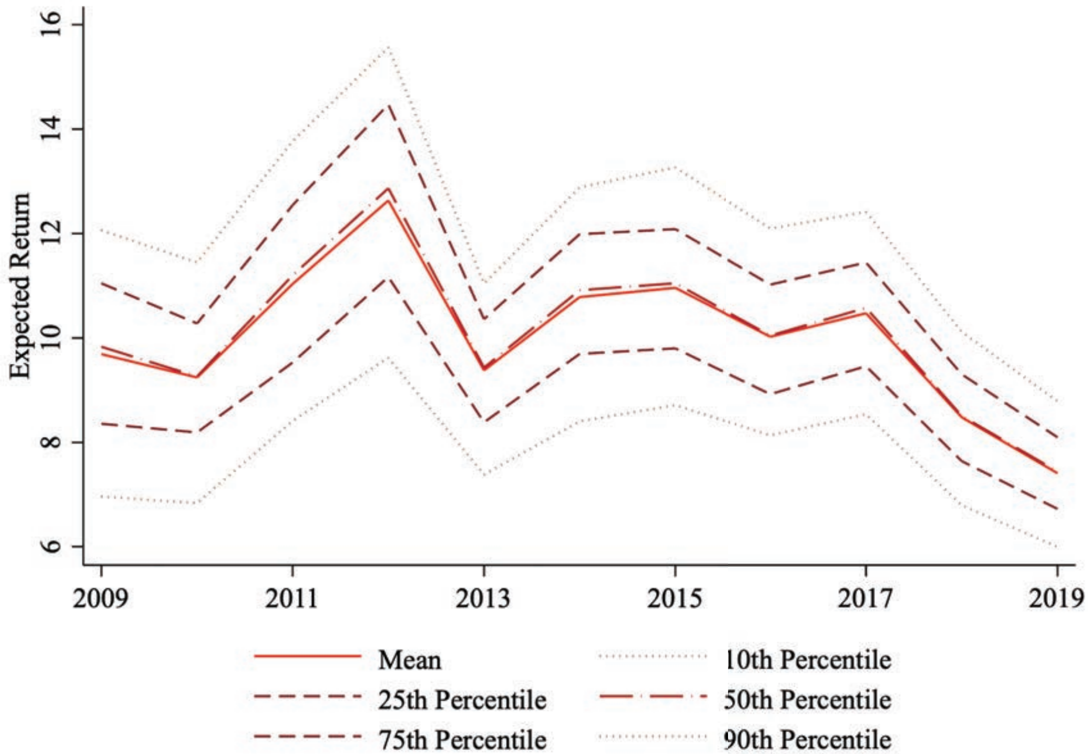
Notes: Figure 5 panel (a) displays the share of assets held in US equities over the period 2009-2019 and the expected share of assets held in US equities if all investors used either a naive 1/N strategy by fund or 1/N strategy by investment category. When computing the share held in US equities, we drop all target date fund assets and assume that remaining non-target date allocation funds hold 60% in US equities. Panels (b) and (c) compare allocations to US equities with the predicted allocations based on the 1/N strategy and allocations not explained by the 1/N strategy, respectively. To predict allocation, we regress observed equity shares on equity shares implied by 1/N strategy by fund, along with year and 2-digit NAICS fixed effects.

Figure 6: Risk Aversion Over Time



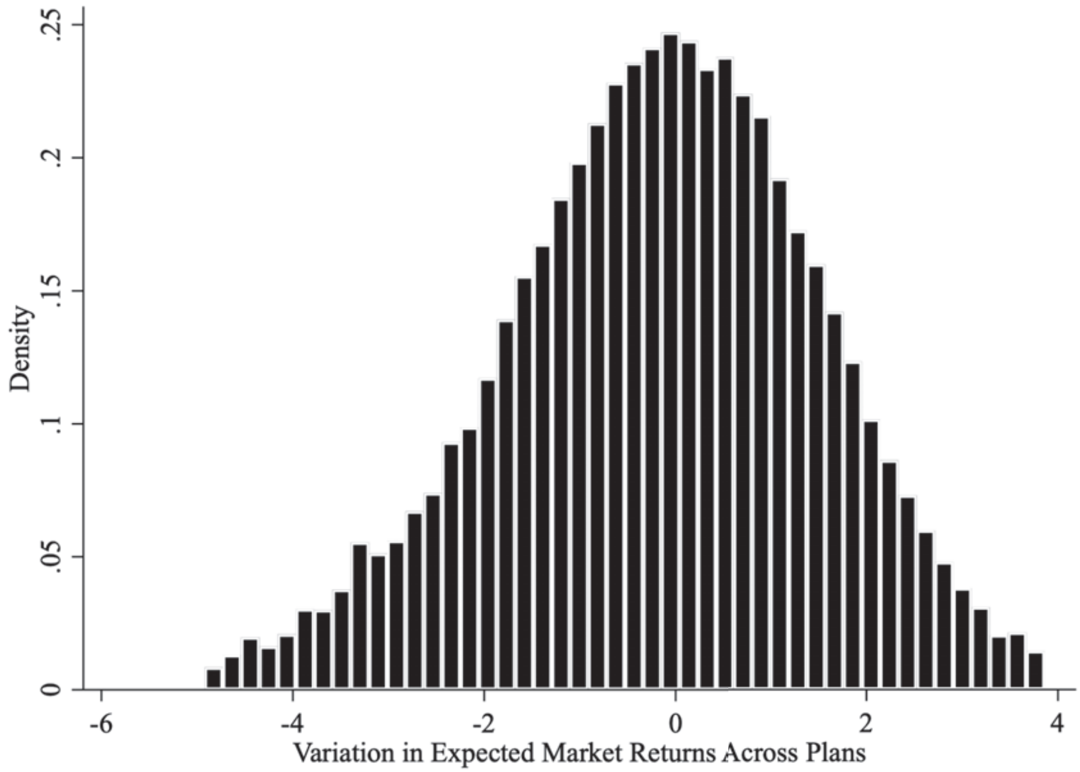
Notes: Figure 6 displays estimated risk aversion over time. Risk aversion corresponds to our model estimates reported in specification (3) of Table 5. When computing risk aversion, we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. Risk aversion is winsorized at the 1% level.

Figure 7: Distribution of Investor Beliefs About the Stock Market Over Time



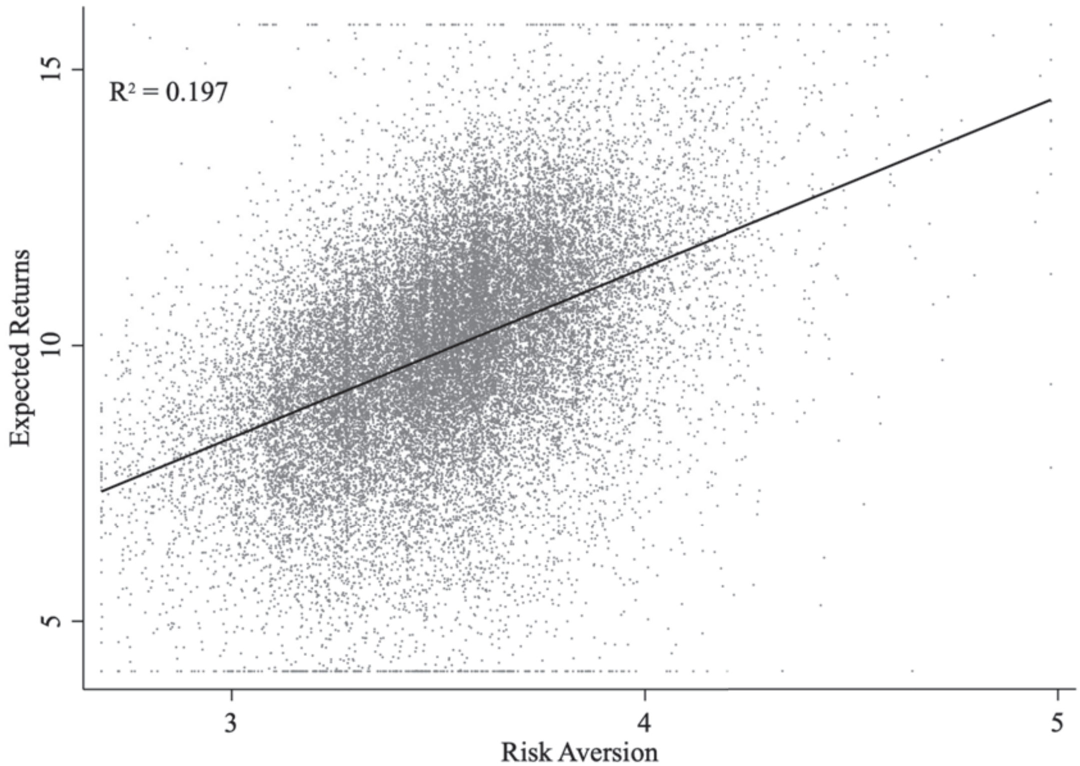
Notes: Figure 7 displays the estimated distribution of investor expectations of market returns. The estimates correspond to the specification reported in specification (3) of Table 5. When computing risk aversion and beliefs, we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. Beliefs are winsorized at the 1% level.

Figure 8: Cross-Section of Investor Beliefs About Stock Market Returns



Notes: Figure 8 displays the estimated cross-sectional distribution of investor expectations of market returns. The estimates correspond to the specification reported in specification (3) of Table 5. When computing risk aversion and beliefs, we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan and winzorize them at the 1% level. Expectations are de-meaned across investors within each year, and each observation reflects the average deviation from the yearly mean over the period 2009-2019. Negative values indicate plans with investors that have persistently pessimistic expectations relative to the mean. Observations are at the plan level. To account for outliers we truncate the distribution at the 1% and 99% percentile.

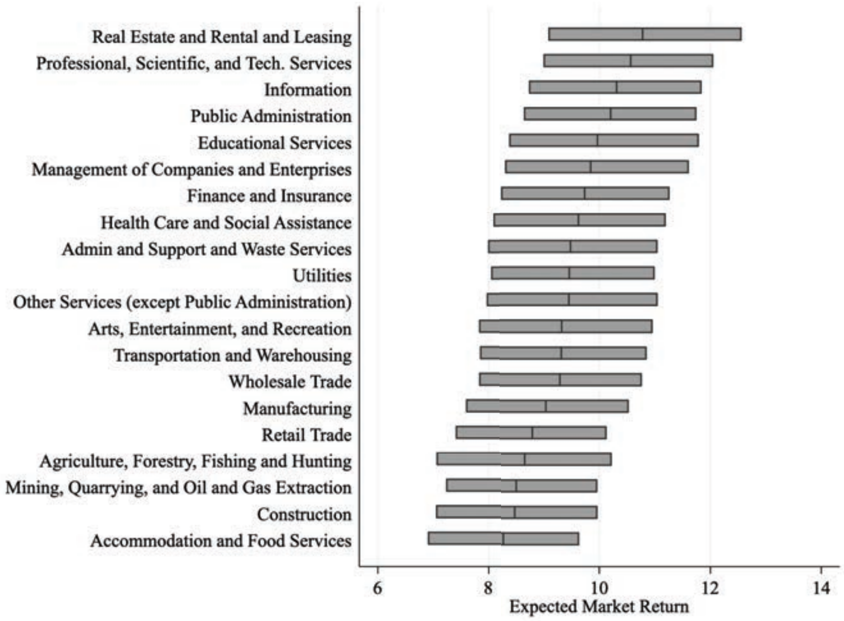
Figure 9: Beliefs About Stock Market Returns vs. Risk Aversion



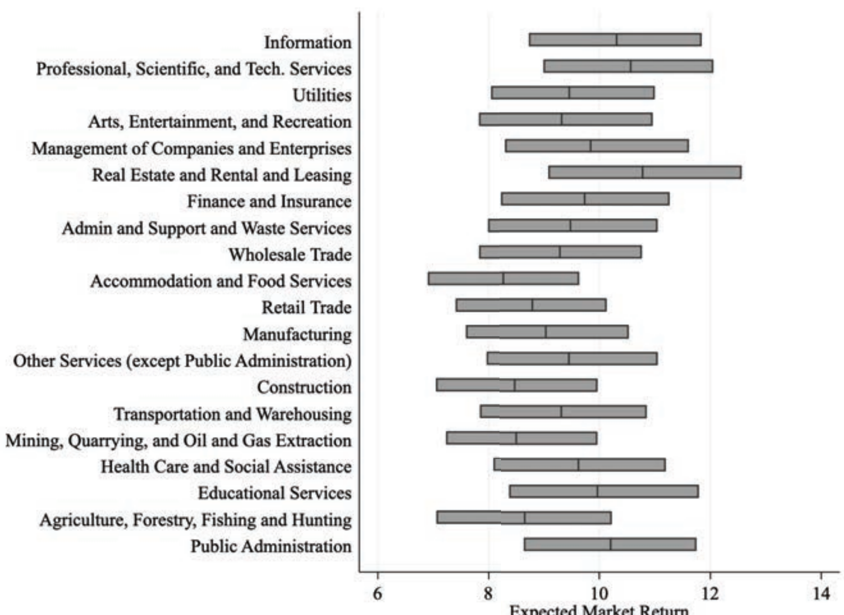
Notes: Figure 9 displays a scatter plot of the cross section of expected returns versus risk aversion as of 2016. The estimates correspond to the specification reported in specification (5) of Table 5. When computing risk aversion and beliefs, we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan and winsorizee them at the 1% level.

Figure 10: Expected Market Returns Across and Within Sectors

(a) Sorted by Median Expected Market Return

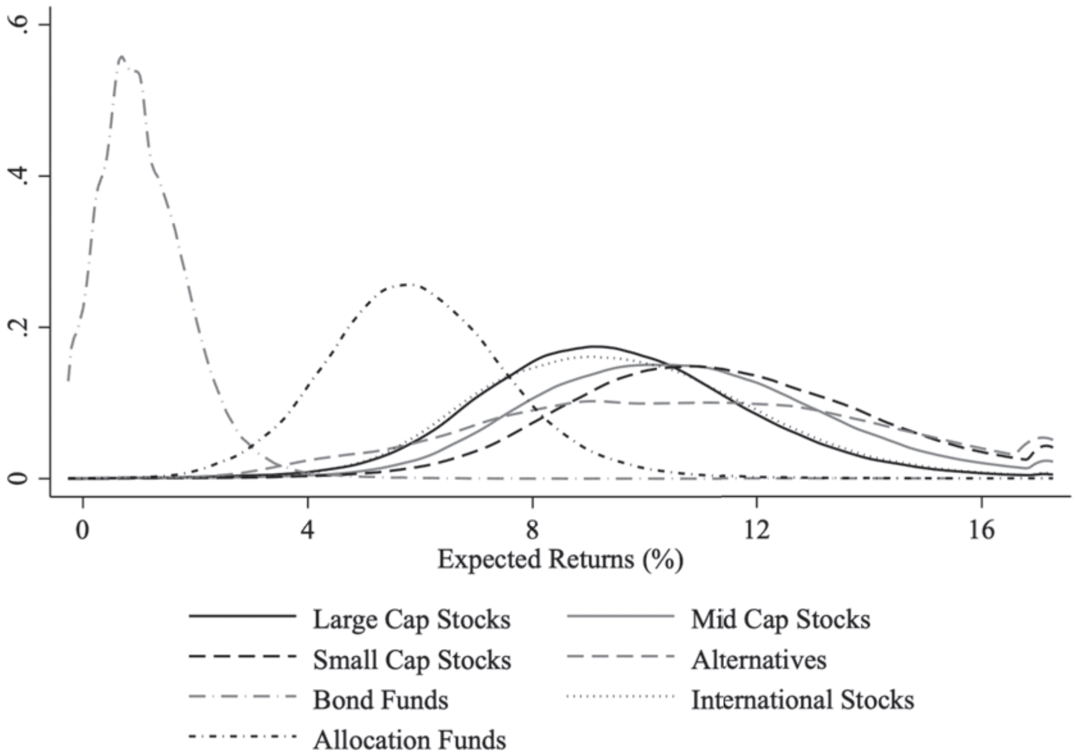


(b) Sorted by Median Equity Allocation



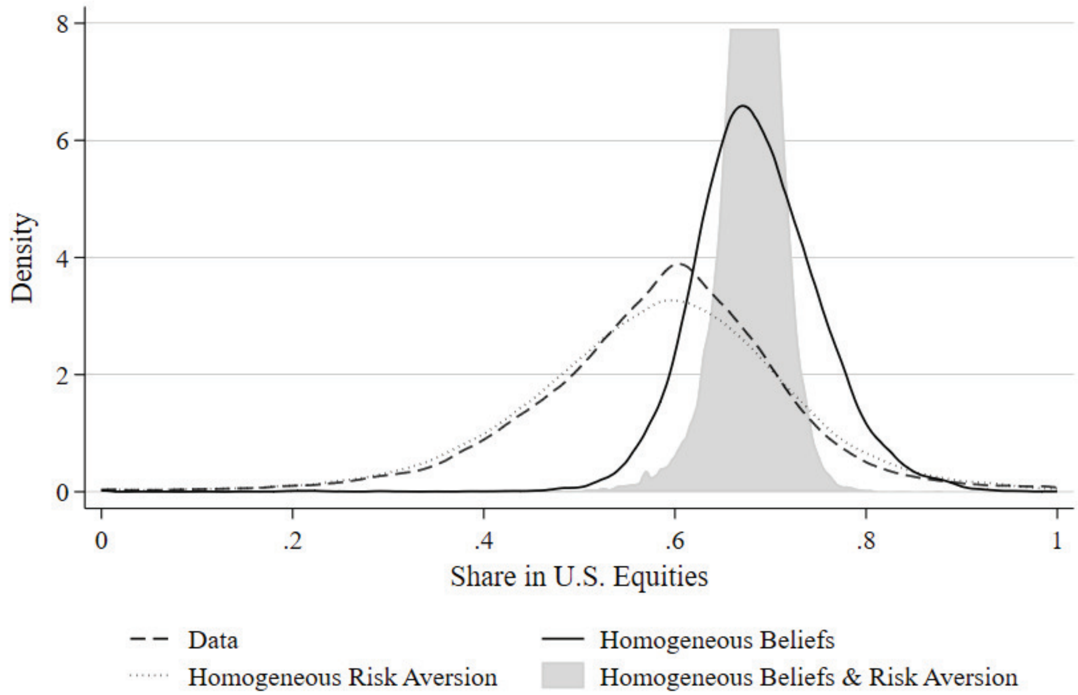
Notes: Figure 10 displays the distribution of expected market returns across sectors (2-digit NAICS). The horizontal gray bars cover the 25th to 75th percentiles, and the short vertical lines indicate medians. Panel (a) is sorted by median expected market return. Panel (b) reports the same data sorted by median U.S. equity allocation (see Figure 3). Expected market returns are calculated based on the specification reported in specification (3) of Table 5 where we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. Observations are at the plan-by-year level over the period 2009-2019.

Figure 11: Distribution of Investor Beliefs by Investment Category



Notes: Figure 11 displays the estimated distributions of investors' expectations of returns across investors for each investment category. The estimates correspond to the specification reported in specification (3) of Table 5. When computing risk aversion and beliefs, we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. For each plan in each year and category, we compute the average expected return at the category level by averaging expected returns across all investment options available in the corresponding category.

Figure 12: Counterfactual Allocations without Heterogeneity in Beliefs or Risk Aversion



Notes: Figure 12 displays actual and counterfactual densities of equity allocations by plan in 2016. The dash line indicates the actual distribution of equity allocations across plans. The solid line indicates the counterfactual (optimal) allocations under the assumption that every investor has identical beliefs about each fund. The dotted line indicates the counterfactual allocations when investors have identical risk aversion parameters. The gray shaded area indicates allocations when investors share identical beliefs and risk aversion. To show all densities on a more reasonable scale, we visually crop the top of this last counterfactual density. When removing heterogeneity for risk aversion and beliefs, we use the mean value of risk aversion across plans weighed by total plan assets, and we use the implied expectations based on aggregate fund balances across plans.

Table 1: Summary Statistics

(a) Plan Summary Statistics

	Obs	Mean	Std. Dev.	Median
Total Assets (millions)	442,631	84.7	689.7	10.7
Number of Plan Participants	425,075	1,261	92,360	223
Number of Investment Options	442,631	26.3	13.8	26.0
Average Account Balance	424,136	66,082	532,846	45,324
Plan Participation Rate	405,832	0.738	0.922	0.833
Employer Contribution Rate	392,401	0.337	0.245	0.290
Share Retired	406,258	0.008	0.014	0.001
Investment Category:				
US Equities	442,631	0.441	0.192	0.455
Target Date Funds	442,631	0.230	0.260	0.137
Bond Fund	442,631	0.126	0.096	0.106
Cash	442,631	0.113	0.127	0.078
International Stock	442,631	0.082	0.072	0.067
Alternatives	442,631	0.009	0.019	0.000
Investment Vehicle Type:				
Mutual Fund	442,631	0.612	0.407	0.823
Separate Account	442,631	0.191	0.356	0.000
Guaranteed Investment Contract	442,631	0.080	0.114	0.038
Collective Trust	442,631	0.053	0.169	0.000
Company Stock	442,631	0.030	0.154	0.000
Common Stock	442,631	0.010	0.076	0.000
Brokerage	442,631	0.009	0.054	0.000
Other	442,631	0.014	0.084	0.000

(b) Investment Option Summary Statistics

	Obs	Mean	Std. Dev.	Median
Volatility	10,781,851	0.137	0.043	0.148
Expense Ratio (pp; BrightScope)	1,856,108	0.569	0.383	0.590
Expense Ratio (pp; CRSP)	6,596,581	0.606	0.432	0.610

Notes: Table 1a displays plan level summary statistics. Observations are reported at the plan-by-year level over the period 2009-2019. Table 1b displays investment option-by-plan-by-year level summary statistics. Observations for *Expense Ratio (BrightScope)* are at the investment option-by-plan level as of 2019. Observations for all other variables are at the investment option-by-plan-by-year level over the period 2009-2019. *Volatility* corresponds to the dependent variable in eq. (5) and is annualized.

Table 2: Equity Allocation vs. Demographics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age	-0.166											-0.171
	(0.119)											(0.127)
ln(Income)		0.643***										-0.029
		(0.164)										(0.198)
ln(Home Value)			0.881***									0.073
			(0.216)									(0.233)
College				1.339***								0.861***
				(0.183)								(0.193)
Employed					0.144*							0.080
					(0.080)							(0.081)
Black						-0.244*						-0.180
						(0.148)						(0.151)
Hispanic							-0.787***					-0.505***
							(0.155)					(0.166)
Unionized								-0.730***				-0.627**
								(0.249)				(0.246)
Sector Equity Beta									0.081**			0.081**
									(0.034)			(0.034)
Share Retired										-0.277***		-0.400***
										(0.056)		(0.055)
ln(Avg. Acct. Bal.)											0.852***	0.892***
											(0.069)	(0.069)
Observations	243,166	243,166	243,166	243,166	243,166	243,166	243,166	243,166	243,166	243,166	243,166	243,166
R^2	0.165	0.166	0.165	0.166	0.165	0.165	0.166	0.165	0.165	0.166	0.169	0.171
Year FE	X	X	X	X	X	X	X	X	X	X	X	X
NAICS FE	X	X	X	X	X	X	X	X	X	X	X	X
County FE	X	X	X	X	X	X	X	X	X	X	X	X

Notes: Table 2 displays the regression results corresponding to a linear regression model. The dependent variable is the share of the 401(k) held in US equities. When computing the share held in US equities, we drop all target date fund assets and assume that remaining non-target date allocation funds hold 60% in US equities. The independent variables, other than the dummy variable *Union*, are all standardized such that they are in units of standard deviations. Observations are at the plan-by-year level over the period 2009-2019. Standard errors are clustered 2-digit NAICs by county level and are in parenthesis. *** p<0.01, ** p<0.05, * p<0.10.

Table 3: Portfolio Allocation vs. Expenses

VARIABLES	(1)	(2)
Expense Ratio	-0.576***	-0.672***
	(0.003)	(0.007)
Observations	5,063,093	5,048,630
Plan×Year FE	X	X
Category×Year×Index FE	X	X
IV		X

Notes: Table 3 displays the regression results corresponding to a linear regression model (eq. 2). Observations are at the investment option-by-plan-by-year level over the period 2009-2019 where we exclude target date funds. The dependent variable is the log share of plan assets held in the investment option. Expense ratios are measured in terms of percentage points. We estimate column (2) using 2-stage least squares. We instrument for expenses using Hausman-type instruments where we instrument for the expenses for a fund using the average expenses of other funds managed by the same fund manager in different Lipper objective categories. Standard errors are clustered at the plan level and are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 4: Autocorrelation in Portfolio Weights

VARIABLES	(1)	(2)
Expected Portfolio Weight	0.767***	0.885***
	(0.006)	(0.008)
Observations	3,737,737	2,875,414
R^2	0.589	0.784
Excluding Newly Added Funds		X

Notes: Table 4 displays the one year autocorrelation in portfolio weights. Observations are at the investment option-by-plan-by-year level over the period 2009-2019 where we restrict our attention to those 401(k)s that were available for at least a year. For ease of interpretation, all dependent and independent variables are standardized such that coefficient estimates are equivalent to correlation coefficients. We compute *Expected Portfolio Weight* under the assumption that the portfolio weight of a fund grows by the return of fund relative to the total return of the 401(k) portfolio over the same period (assuming no rebalancing). Column (1) includes investment options that were not available in the fund menu in the prior year, and hence *Expected Portfolio Weight* is equal to zero for these options. In column (2) we restrict the sample to those investment options that were available in the prior year. Standard errors are clustered at the plan level and are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$..

Table 5: Estimated Model Parameters and Risk Aversion

	(1)		(2)		(3)		(4)		(5)	
	θ	λ	θ	λ	θ	λ	θ	λ	θ	λ
Fee	-0.193***	*5.171*	-0.212***	*4.718*	-0.191***	*5.246*	-0.281***	*3.558*	-0.262***	*3.823*
	(0.017)		(0.016)		(0.030)		(0.028)		(0.037)	
× Age			0.017**	*0.380*	0.018**	*0.500*	0.017**	*0.217*	0.018**	*0.267*
			(0.007)		(0.008)		(0.007)		(0.008)	
× Frac Black			0.005	*0.104*	0.005	*0.125*	0.005	*0.061*	0.005	*0.068*
			(0.005)		(0.005)		(0.005)		(0.005)	
× Frac Hispanic			-0.004	*-0.096*	-0.004	*-0.111*	-0.004	*-0.053*	-0.004	*-0.057*
			(0.011)		(0.011)		(0.011)		(0.011)	
× Frac College			0.025**	*0.556*	0.025**	*0.675*	0.025**	*0.314*	0.024**	*0.356*
			(0.010)		(0.010)		(0.010)		(0.010)	
× ln(Median Family Income)			-0.021*	*-0.465*	-0.022*	*-0.601*	-0.021*	*-0.261*	-0.022*	*-0.316*
			(0.012)		(0.012)		(0.012)		(0.012)	
× ln(Median House Value)			0.009	*0.193*	0.009	*0.242*	0.009	*0.109*	0.009	*0.127*
			(0.011)		(0.011)		(0.011)		(0.011)	
× Frac Employed			-0.005	*-0.118*	-0.004	*-0.103*	-0.005	*-0.068*	-0.004	*-0.055*
			(0.006)		(0.006)		(0.006)		(0.006)	
× Unionized			0.016	*0.352*	0.015	*0.403*	0.016	*0.203*	0.015	*0.218*
			(0.016)		(0.016)		(0.016)		(0.016)	
× Share Retired			-0.003	*-0.068*	-0.003	*-0.082*	-0.003	*-0.040*	-0.003	*-0.046*
			(0.006)		(0.006)		(0.006)		(0.006)	
× ln(Avg. 401(k) Balance)			-0.000	*-0.007*	0.000	*0.005*	-0.001	*-0.008*	-0.000	*-0.002*
			(0.005)		(0.005)		(0.005)		(0.005)	
× Existing 401(k) Plan							0.070***	*0.885*	0.072***	*1.054*
							(0.023)		(0.023)	
Observations	4,932,059		4,528,147		4,528,147		4,528,147		4,528,147	
Year FE	X		X		X		X		X	
Year-Fee Interactions					X				X	
Estimated Risk Aversion										
Mean	5.171		4.781		5.236		3.599		3.553	
Std. Dev.	0.000		0.628		0.730		0.286		0.472	
Median	5.171		4.769		5.266		3.604		3.506	
Observations	442,631		402,497		402,497		243,268		243,268	

Notes: Table 5 displays estimates corresponding to eq. (5). For each specification, the left column (θ) reports the linear regression estimates and standard errors, and the right column translates the coefficients in terms of in terms of risk aversion (λ) and the marginal effects for the average plan in 2009. Observations are at the investment option-by-plan-by-year level over the period 2009-2019. The dependent variable is the additional risk of investing a dollar in a given investment option, given the other portfolio holdings in the plan. The independent variables, other than the dummy variables *Union* and *Existing 401(k) Plan* equal to zero, are all standardized such that they are in units of standard deviations. We estimate each specification using 2-stage least squares. We instrument for expenses and the corresponding interaction terms using Hausman-type instruments as described in the text. We weight each observation by the total assets of the 401(k) plan. All specifications include plan-by-year, Morningstar investment category-by-BrightScope investment category-by-year fixed effects, and index-fund-by-year fixed effects. Standard errors are clustered at the plan level and are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$. The bottom panel reports summary statistics of estimated risk aversion at the plan-by-year level.

Table 6: Residualized Variation in Expected Market Returns vs. Demographics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age	-0.140***											-0.200***
	(0.033)											(0.047)
ln(Income)		0.173**										0.076
		(0.064)										(0.072)
ln(Home Value)			0.154***									-0.002
			(0.048)									(0.041)
College				0.226***								0.114
				(0.073)								(0.083)
Employed					0.123*							0.021
					(0.061)							(0.033)
Black						-0.113***						-0.101***
						(0.037)						(0.025)
Hispanic							-0.101*					-0.083**
							(0.051)					(0.032)
Unionized								-0.553***				-0.412***
								(0.135)				(0.107)
Sector Equity Beta									0.022***			0.022***
									(0.005)			(0.005)
Share Retired										-0.118***		-0.110***
										(0.019)		(0.018)
ln(Avg. Acct. Bal.)											0.099**	0.076
											(0.041)	(0.054)
Observations	243,268	243,268	243,268	243,268	243,268	243,268	243,268	243,268	243,268	243,268	243,268	243,268
R^2	0.006	0.010	0.008	0.017	0.005	0.004	0.003	0.005	0.000	0.004	0.003	0.039
Year FE	X	X	X	X	X	X	X	X	X	X	X	X

Notes: Table 6 displays the regression results corresponding to a linear regression model. Observations are at the investment plan-by-year level over the period 2009 through 2019. The dependent variable is the residual variation in expected market returns, residualized on risk aversion. Expected market returns are calculated based on the specification reported in specification (3) of Table 5 where we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. Both risk aversion and expected market returns are winsorized at the 1% level. Standard errors are clustered 2-digit NAICs level and at the county level and are in parenthesis.. *** p<0.01, ** p<0.05, * p<0.10.

Table 7: Equity Holdings vs. Beliefs and Risk Aversion

VARIABLES	(1) All Equities	(2) All Equities	(3) US Equities	(4) US Equities	(5) Cash	(6) Cash
Risk Aversion (Std.)	-6.511***	-7.449***	-5.702***	-6.303***	4.400***	5.006***
	(0.192)	(0.153)	(0.135)	(0.156)	(0.262)	(0.341)
Expected Returns (Std.)	9.974***	13.692***	8.561***	12.031***	-7.140***	-9.738***
	(0.367)	(0.245)	(0.280)	(0.176)	(0.425)	(0.392)
Observations	243,268	243,268	243,268	243,268	243,268	243,268
R^2	0.507	0.788	0.348	0.595	0.286	0.440
Year FE		X		X		X

Notes: Table 7 displays the regression results corresponding to a linear regression model. Observations are at the plan-by-year level over the period 2009-2019. The dependent variable is the share of the plan portfolio in equities in columns (1) and (2); in US equities in columns (3) and (4); and cash in columns (5) and (6). When computing the shares, we drop all target date fund assets and assume that remaining non-target date allocation funds hold 60% in US equities. Expected market returns and risk aversion are calculated based on the specification reported in specification (3) of Table 5 where we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. Risk aversion and beliefs are both winsorized at the 1% level to account for outliers. Standard errors are clustered 2-digit NAICs level by year level and the plan level and are in parenthesis. *** p<0.01, ** p<0.05, * p<0.10.

Table 8: Expected Returns vs. Past Fund Returns

VARIABLES	(1)	(2)	(3)	(4)
Lag Fund Ret.	0.005***	0.016***	0.007***	0.005***
	(0.000)	(0.001)	(0.002)	(0.000)
Lag Fund Ret. $\times$ New Investment				-0.000
				(0.000)
Observations	4,499,736	672,910	79,041	4,499,736
R^2	0.937	0.941	0.940	0.937
Multi-Level FE	X	X	X	X
New Funds		X		
New Plans			X	

Notes: Table 8 displays the regression results corresponding to a linear regression model. Observations are at the investment option-by-plan-by-year level over the period 2009 through 2019. The dependent variable is the expected returns of the fund. Expected returns are calculated based on the specification reported in specification (3) of Table 5 where we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. Expected returns are winsorized at the 1% level. Each specification include plan-by-year, investment category (Morningstar$\times$BrightScope)-by-year, and index fund-by-year fixed effects. Standard errors are in parenthesis and are clustered at the plan level. *** p<0.01, ** p<0.05, * p<0.10.

Table 9: Expected Market Returns vs. Local Economic Conditions

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Pop. Growth	0.125***	0.041***							0.117***	0.032**
	(0.016)	(0.012)							(0.021)	(0.013)
Home Price Growth			0.022***	0.005***					0.008	0.003
			(0.006)	(0.002)					(0.007)	(0.002)
Establishment Growth					0.039***	0.016***			-0.022*	0.006
					(0.011)	(0.004)			(0.012)	(0.004)
GDP Growth							0.036***	0.005***	0.024***	0.003**
							(0.004)	(0.001)	(0.004)	(0.002)
Observations	232,877	225,188	239,199	231,551	243,268	235,577	239,313	231,731	225,022	217,483
R^2	0.357	0.871	0.344	0.865	0.343	0.864	0.344	0.864	0.359	0.872
Year FE	X	X	X	X	X	X	X	X	X	X
Plan FE		X		X		X		X		X

Notes: Table 9 displays the regression results corresponding to a linear regression model. Observations are at the investment plan-by-year level over the period 2009 through 2019. The dependent variable is the expected return of the market. The expected return of the market is calculated based on the specification reported in specification (3) of Table 5 where we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. Expected returns are winsorized at the 1% level. Standard errors are in parenthesis and are clustered at the county-by-year level. *** p<0.01, ** p<0.05, * p<0.10.

Table 10: Expected Market Returns vs. Employer Returns

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Firm Return (1 years)	0.001**	0.001***							0.001**	0.001***
	(0.000)	(0.000)							(0.000)	(0.000)
Firm Investment			0.004**	0.018***					0.005**	0.016***
			(0.002)	(0.003)					(0.002)	(0.004)
Sales Growth					0.000	0.004***			-0.000	0.001**
					(0.000)	(0.001)			(0.000)	(0.001)
Employment Growth							0.001	0.005***	0.000	0.002*
							(0.001)	(0.001)	(0.001)	(0.001)
Observations	11,495	11,738	10,262	10,474	11,233	11,452	11,216	11,441	9,889	10,081
R^2	0.886	0.510	0.889	0.521	0.886	0.510	0.887	0.510	0.890	0.519
Year FE	X		X		X		X		X	
Plan FE	X		X		X		X		X	
NAICS×Year FE		X		X		X		X		X

Notes: Table 10 displays the regression results corresponding to a linear regression model. Observations are at the investment plan-by-year level over the period 2009 through 2019. The dependent variable is the expected return of the market. The expected return of the market is calculated based on the specification reported in specification (3) of Table 5 where we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. Expected returns are winsorized at the 1% level. Standard errors are in parenthesis and are clustered at the plan level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 11: Predictability of Forecast Errors

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Lag Forecast Error	0.035***	0.002				
	(0.001)	(0.001)				
Lag Fund Ret.			-0.030***	-0.029***		
			(0.001)	(0.001)		
Change in Beliefs					-0.511***	-0.795***
					(0.007)	(0.012)
Observations	2,400,158	2,395,689	4,494,924	4,494,868	2,402,780	2,398,321
R^2	0.627	0.662	0.616	0.648	0.628	0.664
Year FE	X		X		X	
Plan×Year FE		X		X		X

Notes: Table 11 displays the regression results corresponding to a linear regression model (eq. 16). Observations are at the investment option-by-plan-by-year level over the period 2009 through 2019. The dependent variable is investor's forecast errors as measured per eq. (15). *Lag Forecast Error* measures investors forecast error in the previous period. *Lag Fund Ret.* measures the annual fund return in the previous year. *Change in Beliefs* measures the change in investor's beliefs about the expected returns of the fund over the previous year. Standard errors are clustered at the plan level. *** p<0.01, ** p<0.05, * p<0.10.

A Additional Analysis and Robustness

A.1 Allocations and Investor Characteristics

Here, we present additional details about the reduced-form relationships between investor characteristics and equity allocations, as well as examining the relationships for other asset classes.

Wealth and Income: Plans with wealthier participants, measured by average account balances, allocate more towards equities. The results in column (12) indicate that a one standard deviation increase in the average account balance is correlated with a 0.89 pp increase in equity exposure. Previous research based on data from the Survey of Consumer Finances in the US (Heaton and Lucas, 2000; Campbell, 2006; Wachter and Yogo, 2010) and administrative data in Sweden and Norway (Bach et al., 2020; Fagereng et al., 2020) document a similar positive relationship between wealth and equity allocation. Because we study 401(k) portfolios conditional on participation, our results indicate that the positive relationship between wealth and equity allocation is not solely driven by participation costs along the extensive margin.

Similarly, we find that income and home wealth are positively correlated with equity exposure, although the effect becomes insignificant due to multicollinearity once we include other controls in column (12). The existing theoretical predictions regarding how equity exposure vary by income and home value are mixed.[42]

Age and Retirement: We find that age and share of retired participants are negatively correlated with equity exposure. One standard deviation increases in participant age and the share of participants retired are associated with a 0.17 and 0.40 pp decline in US equity holdings, respectively (column 12). The decreasing age profile is consistent with standard life cycle models (Cocco et al., 2005) which consider the present value of future income as safe assets. Using novel survey data, Choi and Robertson (2020) find that years left until retirement is one of the most commonly cited factors for determining equity allocations.[43]

Other Demographics: We also find that more educated households have higher equity allocation. The results in column (4) indicate that a one standard deviation increase in the share

[42] Cocco et al. (2005) shows how income is analogous to a safe asset, and hence is positively correlated investment in risky equity; however, other theoretical works highlight how income risk can also crowd out equity allocation (Lynch and Tan, 2011; Storesletten et al., 2004). Housing can also be considered as a long-term safe asset and hedges against rental prices (Sinai and Souleles, 2005). Housing also provides collateral for borrowing, and can increase equity holding thanks to lower borrowing constraints (Guiso et al., 1996). On the other hand, housing is illiquid. In life cycle models with housing decisions, Cocco (2005) and Yao and Zhang (2005) show that individuals with a higher fraction of total wealth in real estate invest less in risky assets.

[43] Empirical estimates tend to be mixed due to the identification challenge of collinearity among cohort, time and age effect. Using Norwegian administrative data, Fagereng et al. (2017) find that risky asset share of stock market participants is a decreasing function of age. However, **?** and Ameriks and Zeldes (2004) find evidence of hump-shaped patterns based on US data.

of college educated individuals is correlated with a 1.34 pp increase in equity allocation This relationship is consistent with the findings in Campbell (2006) and Black et al. (2018), and could potentially be driven by financial literacy (Calvet et al., 2007; Van Rooij et al., 2011).

We find that minorities invest less in equity. A one standard deviation increase in the fraction of Hispanic and black populations are correlated with 0.79 and 0.24 percentage point decreases in equity exposure. Campbell (2006) and Chiteji and Stafford (2000) also find that minorities have lower equity shares.

Other Asset Classes

The differences in equity allocation across plans documented in Section 3.1 extends to other asset classes as well. Appendix Table A2 displays the regression results where we replicate eq. (1) for the other main asset classes. The dependent variable in columns (1)-(2) is the portfolio share in US equities, in columns (3)-(4) is the share in bonds, in columns (5)-(6) is the share in cash, and in columns (7)-(8) is the share in international equities. A couple of interesting patterns emerge in Table A2. In general, the demographics that are positively (negatively) correlated with US equity ownership are also positively (negatively) correlated with international equity ownership with a few notable exceptions. For example, education is positively correlated with both US equity ownership and international equity ownership. However, wealth, as measured by account balances, is positively correlated with US equity ownership but negatively correlated with international equity ownership. These findings regarding international exposure are consistent with the evidence in Bekaert et al. (2017). Plans with a greater share of retirees and older participants tend to have higher bond and cash exposures and lower US and international equity exposures. Union membership and minority status are correlated with higher cash allocations but are negatively correlated with equity and bond allocations.

In Appendix Table A4, we replicate Appendix Table A2 where we control for the composition of the 401(k) menu. We find that controlling for the composition of the menu has little impact on our estimates. We also show in Appendix Table A3 that the menus themselves are largely uncorrelated with participant demographics.

A.2 Relation to Future Returns

We examine the relationship between investor expectations and return in the following regression:

$$Ret_{kt'} = \bar{\mu}_{kt} + \eta_{kt.} \tag{17}$$

Observations are at the fund-by-year level. The dependent variable is the return of fund k over the period t to t', where we examine the forecastability of returns over a year horizon. We control for the mean expected return of fund k at time t across plans ($\bar{\mu}_{kt}$) and the interquartile range of expected returns of fund k at time t across plans.

Table A13 displays the corresponding estimates. We examine the predictability of returns over a one year horizon in columns (1) and (3) and over a three year horizon in columns (2) and (4). In columns (1) and (2) we do not control for fund risk, while in columns (3) and (4) we control for fund risk as measured by the fund's factor loadings.[44] Controlling for fund risk is important because otherwise investor expectations could just be capturing differences in fund risk. Consistent with this intuition, we find a positive and significant relationship between investor expectations and future returns in columns (1) and (2) when we omit risk controls. However, once we control for differences in risk in columns (3) and (4), , the relationship between investor expectations and future returns disappears. Thus, investor expectations do not forecast future returns once we account for known differences in risk.

A.3 Accounting for Labor Income Risk

We also consider the case when investors account for labor income risk. Specifically, we model an investor's labor income risk as an additional asset with a fixed relative weight ϖ (relative to the value of the investor's 401(k) portfolio) and factor loadings b_{wlt} for each factor l. We can then rewrite an investor's first order condition as:

$$\mu_{ikt} - p_{kt} - R_F = \lambda \left(\sum_{l=1}^{L} b_{klt} \left(b_{wlt}\varpi + \sum_{j=1}^{K} b_{jlt}\omega_{ijt} \right) + \sigma_k^2 \omega_{ikt} \right).$$

Rearranging the terms yields:

$$\left(\sum_{l=1}^{L} b_{klt} \left(\sum_{j=1}^{K} b_{jlt}\omega_{ijt} \right) + \sigma_k^2 \omega_{ikt} \right) = \theta p_{kt} + \psi \left(\sum_{l=1}^{L} b_{klt} b_{wlt} \right) + \epsilon_{kt}, \tag{18}$$

where the parameter θ is the negative inverse of risk aversion (i.e., $\theta = \frac{-1}{\lambda}$), ϵ_{kt} is equal to average investor beliefs divided by risk aversion (i.e., $\epsilon_{kt} = (\bar{\mu}_{kt}^{(m)} - R_F)/\lambda$), and ψ is equal to $-\varpi$.

We estimate the empirical equivalent of eq. (18) as

$$\varsigma_{mkt}^2 = \theta p_{mkt} + \psi \xi_{mkt}^2 + \phi_{mt} + \phi_{j(k)t} + \epsilon_{mkt}, \tag{19}$$

where:

$$\varsigma_{mkt}^2 = \left(\sum_{l=1}^{L} b_{klt} \left(\sum_{j=1}^{K_i} b_{jlt}\bar{\omega}_{jt}^{(m)} \right) + \sigma_k^2 \bar{\omega}_{kt}^{(m)} \right),$$

[44] Specifically, we control for the time-varying factor loadings the 55 factors used to calculate portfolio risk. We also allow coefficients on the factor loadings to vary over time.

and

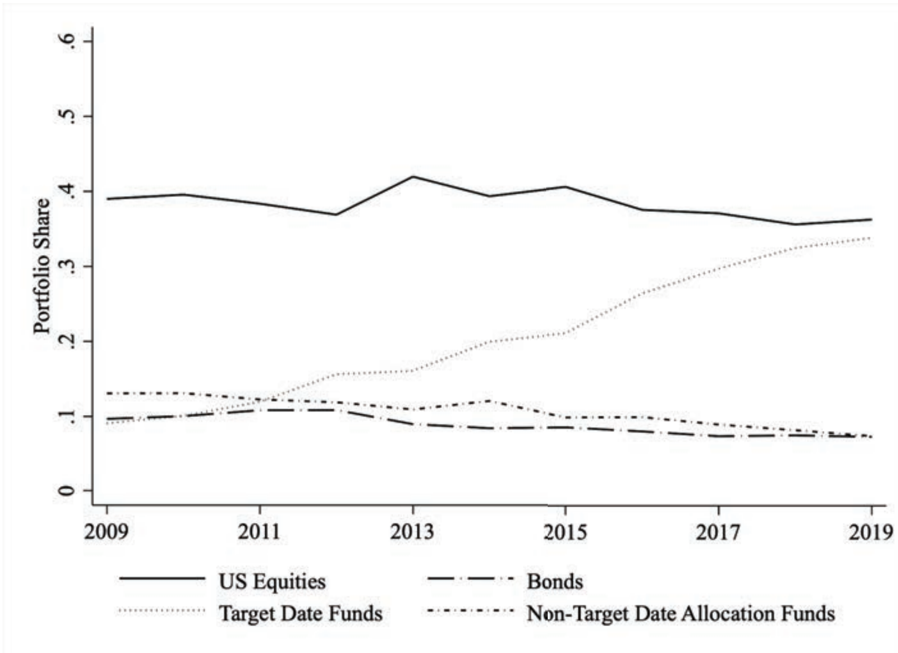
$$\xi_{mkt}^2 = \left(\sum_{l=1}^{L} b_{klt} b_{wlt} \bar{\omega}_{jt}^{(m)} \right).$$

The term ξ_{mkt}^2 captures the additional risk of investing in asset k due to labor income risk. We proxy for the factor loadings for labor income risk using the equity factor loadings corresponding to the industry of the plan sponsor m.

Table A14 displays the corresponding estimates. We estimate a similar inverse risk aversion coefficient θ as in our baseline specification in column (1). In column (2) we include fund-by-year fixed effects, which absorbs the term θ. The object of interest is the parameter $\psi = -\varpi$. Note that in column (1) we estimate $\psi > 0$ which implies a negative weight ϖ such that investors behave as if they are risk seeking with respect to their labor income risk. In column (2) we fail to reject the null hypothesis that $\psi = 0$ such that investors neglect income risk. One caveat is that the additional risk due to labor income $\xi_{mkt}^2 = \left(\sum_{l=1}^{L} b_{klt} b_{wlt} \bar{\omega}_{jt}^{(m)} \right)$ could be correlated with investor beliefs μ, which would make it endogenous in eq. (19). Directly addressing this endogeneity issue is challenging because it requires variation in the additional risk due to labor income that is orthogonal to investor beliefs.

B Additional Tables and Figures

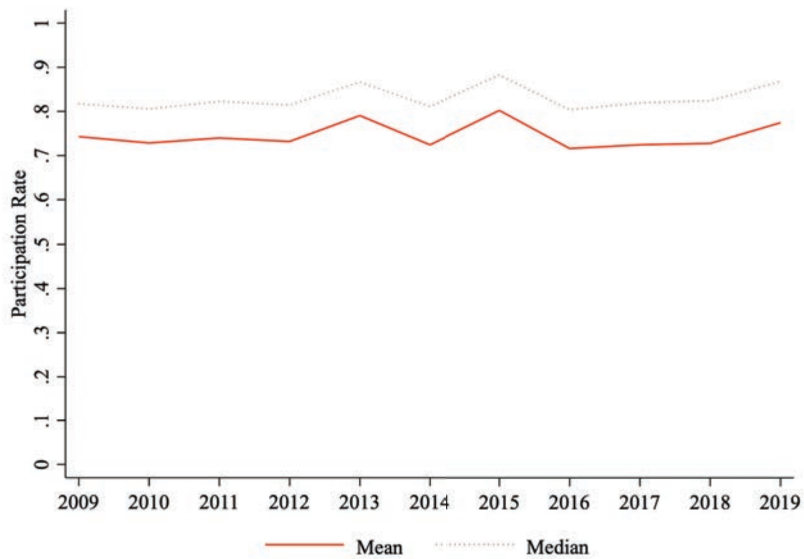
Figure A1: Holdings Over Time



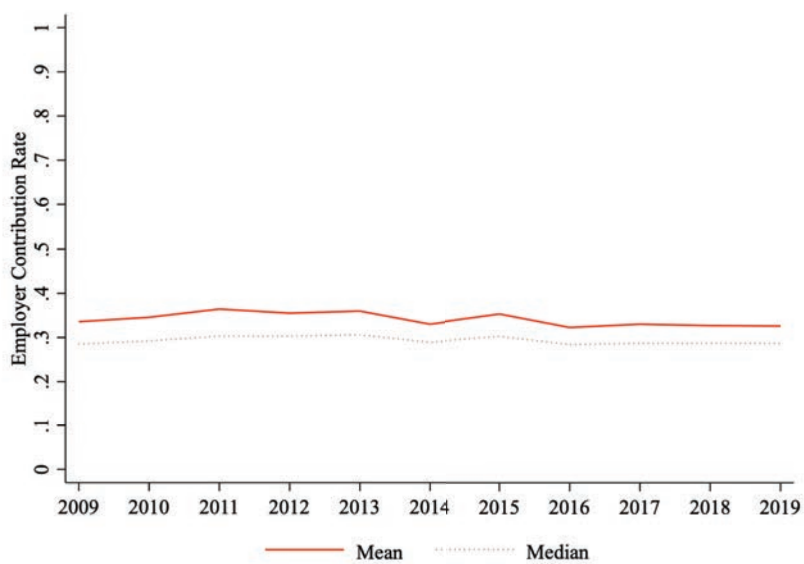
Notes: Figure A1 displays the equal-weighted average holdings of target date and non target date allocation funds, as well as US equity and bond assets without considering allocation funds across plans over the period 2009-2019.

Figure A2: Participation and Employer Contributions Over Time

(a) Participation Over Time

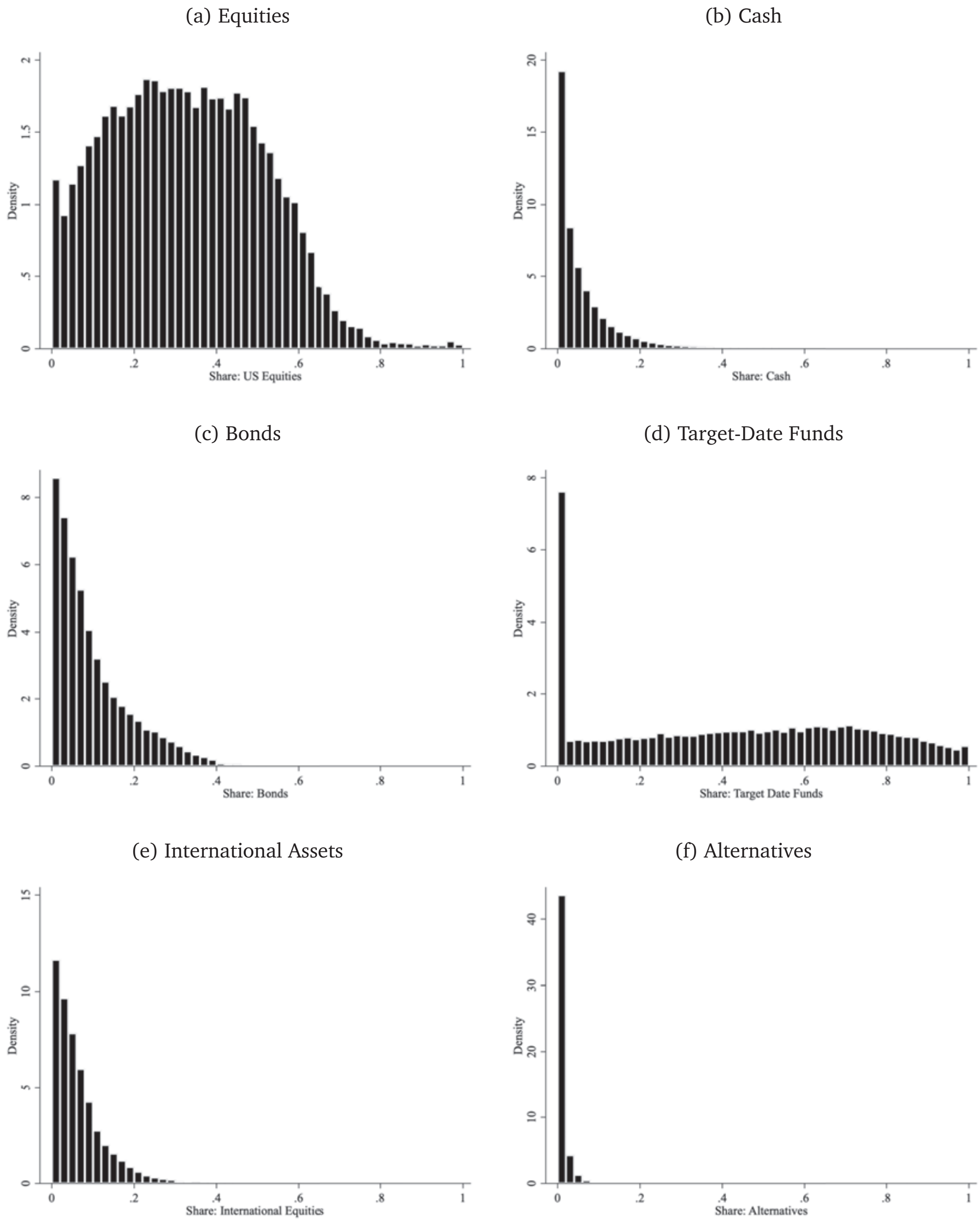


(b) Employer Contributions (Share of Total Contributions)



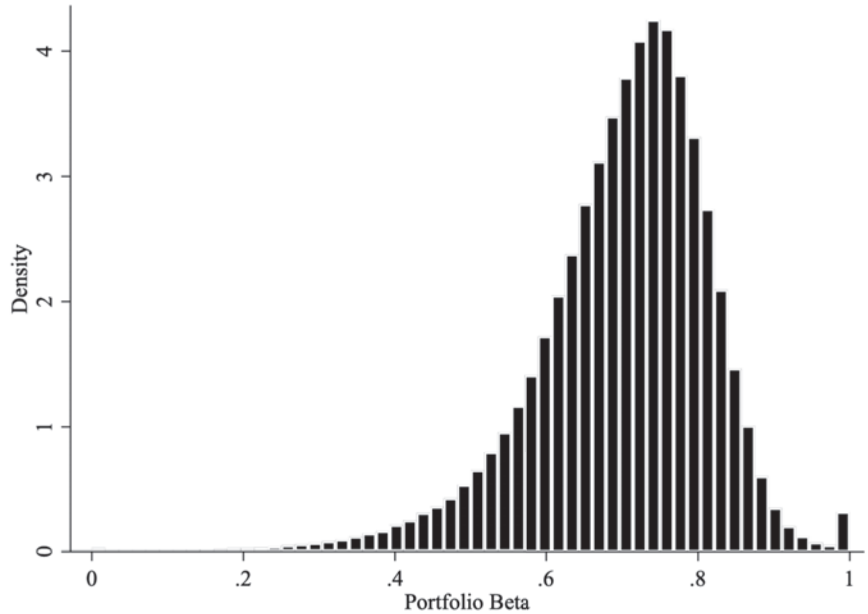
Notes: Figure A2 panel (a) displays the average and median 401(k) participation rate across 401(k) plans over the period 2009-2019. We measure the participation rate as the share of individuals who participate in the plan relative to the number of individuals who are eligible to participate. Panel (b) displays the average and median employer contribution rate across 401(k) plans. The employer contribution rate is measured as the employer's 401(k) contribution relative to the total 401(k) contribution (i.e., employer contribution plus employee contribution).

Figure A3: Distribution of Holdings: Plans Started After 2007



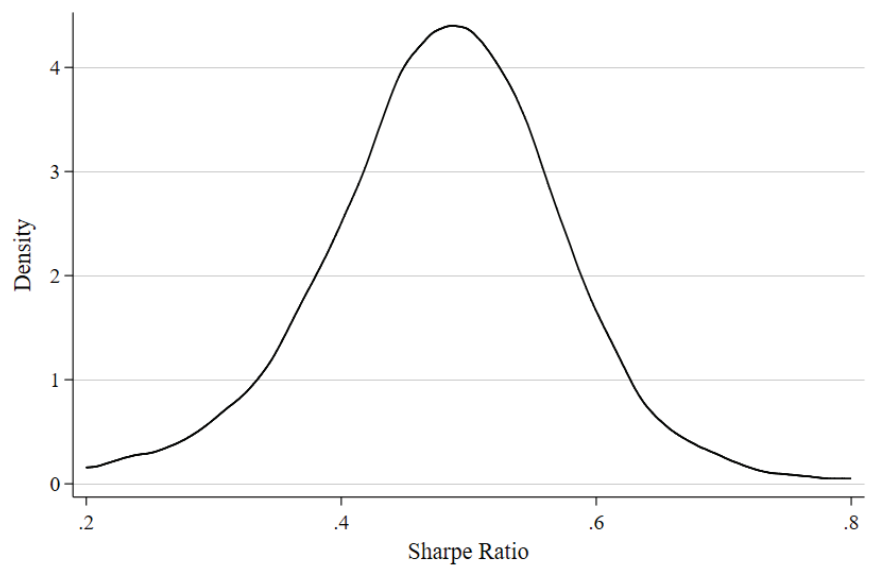
Notes: Figure A3 displays the distribution of holdings across 401(k) plans. Observations are at the plan-by-year level over the period 2009-2019 for those plans with at least five investment options. We also restrict our attention to plans that were started in 2008 or later after the Department of Labor changed the rules for qualified default investment alternatives.

Figure A4: Market Exposure by Portfolio: Equity Beta



Notes: Figure A4 displays the distribution of average equity beta across 401(k) portfolios. Observations are at the plan-by-year level. We compute the average equity beta for a 401(k) plan as the dollar weighted average equity beta across each fund available in the plan. For scaling purposes we truncate the distribution of equity betas at 0 and 1.

Figure A5: Portfolio Sharpe Ratios



Notes: Figure A5 displays the density of implied Sharpe ratios based on plan-level idiosyncratic expected returns and portfolio allocations in 2016.

Table A1: 401(k) Participation vs. Demographics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age	0.011***											0.011***
	(0.003)											(0.003)
ln(Income)		0.017***										-0.007
		(0.003)										(0.006)
ln(Home Value)			0.043***									0.029
			(0.014)									(0.018)
College				0.029***								0.009
				(0.005)								(0.008)
Employed					0.010***							0.009**
					(0.003)							(0.004)
Black						-0.011**						0.000
						(0.005)						(0.005)
Hispanic							-0.009**					-0.000
							(0.004)					(0.004)
Unionized								0.017**				0.020***
								(0.007)				(0.008)
Sector Equity Beta									-0.000**			-0.000**
									(0.000)			(0.000)
Share Retired										0.002		-0.011
										(0.006)		(0.007)
ln(Avg. Acct. Bal.)											0.079***	0.080***
											(0.001)	(0.002)
Observations	242,619	242,619	242,619	242,619	242,619	242,619	242,619	242,619	242,619	242,619	242,619	242,619
R^2	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.036	0.036
Year FE	X	X	X	X	X	X	X	X	X	X	X	X
NAICS FE	X	X	X	X	X	X	X	X	X	X	X	X
County FE	X	X	X	X	X	X	X	X	X	X	X	X

Notes: Table A1 displays the regression results corresponding to a linear regression model. Observations are at the plan-by-year level over the period 2009-2019. The dependent variable is fraction of eligible employees that participate in 401(k) plans. Standard errors are clustered 2-digit NAICs by county level and are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A2: Asset Allocation vs. Demographics

VARIABLES	(1) US Equities	(2) US Equities	(3) Bonds	(4) Bonds	(5) Cash	(6) Cash	(7) Intl. Equities	(8) Intl. Equities
Age	-0.711***	-0.171	0.213***	-0.097	0.848***	0.581***	-0.331***	-0.264***
	(0.119)	(0.127)	(0.048)	(0.072)	(0.103)	(0.110)	(0.047)	(0.082)
ln(Income)	0.458***	-0.029	-0.695***	-0.027	0.476***	-0.087	-0.213***	0.107
	(0.149)	(0.198)	(0.088)	(0.105)	(0.159)	(0.189)	(0.077)	(0.123)
ln(Home Value)	0.157	0.073	-0.400***	-0.339**	0.306**	0.363*	-0.108	-0.103
	(0.115)	(0.233)	(0.068)	(0.151)	(0.127)	(0.207)	(0.067)	(0.149)
College	0.694***	0.861***	0.405***	-0.220*	-1.453***	-0.813***	0.232***	0.184
	(0.113)	(0.193)	(0.068)	(0.130)	(0.106)	(0.185)	(0.064)	(0.130)
Employed	0.137*	0.080	-0.127**	-0.079	0.036	-0.003	-0.011	0.014
	(0.079)	(0.081)	(0.051)	(0.054)	(0.075)	(0.078)	(0.050)	(0.054)
Black	-0.200***	-0.180	-0.111**	-0.073	0.808***	0.046	-0.426***	0.181*
	(0.070)	(0.151)	(0.049)	(0.101)	(0.073)	(0.146)	(0.044)	(0.094)
Hispanic	-0.615***	-0.505***	-0.088	-0.048	0.905***	0.585***	-0.268***	-0.018
	(0.090)	(0.166)	(0.061)	(0.113)	(0.099)	(0.160)	(0.062)	(0.109)
Unionized	-0.407*	-0.627**	-0.675***	-0.439**	3.710***	3.691***	-2.317***	-2.327***
	(0.244)	(0.246)	(0.177)	(0.174)	(0.275)	(0.271)	(0.137)	(0.143)
Sector Equity Beta	0.082**	0.081**	-0.015	-0.013	0.006	0.004	-0.030***	-0.028***
	(0.033)	(0.034)	(0.012)	(0.011)	(0.019)	(0.020)	(0.004)	(0.004)
Share Retired	-0.466***	-0.400***	0.188***	0.134***	0.681***	0.644***	-0.342***	-0.318***
	(0.058)	(0.055)	(0.036)	(0.034)	(0.059)	(0.056)	(0.033)	(0.033)
ln(Avg. Acct. Bal.)	1.058***	0.892***	-0.298***	-0.147***	0.097	0.100	-0.726***	-0.723***
	(0.072)	(0.069)	(0.057)	(0.057)	(0.065)	(0.062)	(0.043)	(0.043)
Observations	243,268	243,166	243,268	243,166	243,268	243,166	243,268	243,166
R^2	0.110	0.171	0.031	0.099	0.077	0.149	0.031	0.096
Year FE	X	X	X	X	X	X	X	X
NAICS FE		X		X		X		X
County FE		X		X		X		X

Notes: Table A2 displays the regression results corresponding to a linear regression model. The dependent variable is the portfolio weight of the corresponding asset class. The independent variables, other than the dummy variable *Union*, are all standardized such that they are in units of standard deviations. Observations are at the plan-by-year level over the period 2009-2019. Standard errors are clustered 2-digit NAICs by county level and are in parenthesis. *** p<0.01, ** p<0.05, * p<0.10.

Table A3: 401(k) Menus vs. Demographics

VARIABLES	(1) US Equity Funds	(2) Bond Funds	(3) Cash Funds	(4) Intl. Equity Funds
Age	0.004***	-0.002***	-0.000	-0.001*
	(0.001)	(0.001)	(0.000)	(0.001)
ln(Income)	-0.001	0.002**	-0.001*	0.000
	(0.001)	(0.001)	(0.000)	(0.001)
ln(Home Value)	0.001	-0.000	0.000	-0.001
	(0.002)	(0.001)	(0.001)	(0.001)
College	0.001	-0.002	0.000	0.001
	(0.002)	(0.001)	(0.001)	(0.001)
Employed	-0.000	-0.000	0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.000)
Black	-0.001	0.000	-0.001	0.001
	(0.001)	(0.001)	(0.000)	(0.001)
Hispanic	0.000	-0.000	-0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Unionized	0.010***	-0.004***	0.009***	-0.012***
	(0.002)	(0.001)	(0.001)	(0.001)
Sector Equity Beta	0.001***	0.000***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Share Retired	-0.004***	0.001***	0.004***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
ln(Avg. Acct. Bal.)	0.003***	-0.002***	0.004***	-0.004***
	(0.001)	(0.000)	(0.000)	(0.000)
Observations	243,166	243,166	243,166	243,166
R^2	0.088	0.067	0.075	0.081
Year FE	X	X	X	X
NAICS FE	X	X	X	X
County FE	X	X	X	X

Notes: Table A3 displays the regression results corresponding to a linear regression model (eq. 10). Observations are at the plan-by-year level over the period 2009-2019. The dependent variable is the number of funds available in the 401(k) menu in a given asset class (e.g., US equities) divided by the total number of funds available in the 401(k) menu. Standard errors are clustered 2-digit NAICs by county level and are in parenthesis. *** p<0.01, ** p<0.05, * p<0.10.

Table A4: Asset Allocation vs. Demographics

VARIABLES	(1) US Equities	(2) Bonds	(3) Cash	(4) International Equities
Age	-0.345***	0.001	0.584***	-0.187***
	(0.117)	(0.068)	(0.109)	(0.066)
ln(Income)	0.028	-0.105	-0.042	0.097
	(0.184)	(0.095)	(0.191)	(0.096)
ln(Home Value)	0.031	-0.323**	0.350*	-0.039
	(0.217)	(0.137)	(0.206)	(0.121)
College	0.825***	-0.148	-0.823***	0.101
	(0.185)	(0.122)	(0.183)	(0.109)
Employed	0.085	-0.078	-0.016	-0.006
	(0.077)	(0.051)	(0.077)	(0.045)
Black	-0.147	-0.083	0.074	0.131*
	(0.139)	(0.094)	(0.144)	(0.072)
Hispanic	-0.513***	-0.038	0.609***	-0.052
	(0.157)	(0.103)	(0.161)	(0.088)
Unionized	-1.054***	-0.269*	3.236***	-1.598***
	(0.240)	(0.161)	(0.266)	(0.118)
Sector Equity Beta	0.059*	-0.016	0.002	-0.009**
	(0.032)	(0.010)	(0.020)	(0.004)
Share Retired	-0.243***	0.077**	0.457***	-0.295***
	(0.054)	(0.032)	(0.054)	(0.027)
ln(Avg. Acct. Bal.)	0.772***	-0.054	-0.110*	-0.459***
	(0.063)	(0.053)	(0.061)	(0.037)
Equity Funds	4.831***			
	(0.066)			
Bond Funds		3.316***		
		(0.044)		
Cash Funds			5.288***	
			(0.127)	
International Funds				4.567***
				(0.049)
Observations	243,166	243,166	243,166	243,166
R^2	0.263	0.200	0.176	0.355
Year FE	X	X	X	X
NAICS FE	X	X	X	X
County FE	X	X	X	X

Notes: Table A4 displays the regression results corresponding to a linear regression model. The dependent variable is the portfolio weight of the corresponding asset class. The independent variables *US Equity Funds*, *Cash Funds*, *Bond Funds*, and *International Funds* are the number of funds available in the 401(k) menu in a given asset class (e.g., US equities) divided by the total number of funds available in the 401(k) menu. The independent variables, other than the dummy variable *Union*, are all standardized such that they are in units of standard deviations. Observations are at the plan-by-year level over the period 2009-2019. Standard errors are clustered 2-digit NAICs by county level and are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A5: Smaller Plans - Asset Allocation vs. Demographics

VARIABLES	(1) US Equities	(2) US Equities	(3) Bonds	(4) Bonds	(5) Cash	(6) Cash	(7) Intl. Equities	(8) Intl. Equities
Age	-0.699***	-0.179	0.303***	0.006	0.859***	0.488***	-0.449***	-0.276**
	(0.131)	(0.157)	(0.066)	(0.100)	(0.117)	(0.138)	(0.065)	(0.109)
ln(Income)	0.517***	0.264	-0.782***	-0.002	0.462***	-0.503**	-0.208**	0.177
	(0.168)	(0.242)	(0.112)	(0.147)	(0.175)	(0.225)	(0.106)	(0.170)
ln(Home Value)	0.132	-0.055	-0.414***	-0.386*	0.397***	0.598**	-0.144	-0.272
	(0.125)	(0.306)	(0.092)	(0.219)	(0.144)	(0.283)	(0.092)	(0.226)
College	0.805***	0.827***	0.397***	-0.352*	-1.417***	-0.406	0.122	-0.068
	(0.136)	(0.269)	(0.090)	(0.185)	(0.125)	(0.254)	(0.087)	(0.197)
Employed	0.018	0.133	-0.066	0.013	0.006	-0.169	0.067	0.034
	(0.101)	(0.113)	(0.068)	(0.079)	(0.092)	(0.106)	(0.068)	(0.081)
Black	-0.191**	-0.227	0.025	0.064	0.762***	0.014	-0.526***	0.099
	(0.093)	(0.204)	(0.069)	(0.142)	(0.096)	(0.194)	(0.059)	(0.133)
Hispanic	-0.562***	-0.547**	-0.075	-0.011	0.890***	0.444**	-0.324***	0.072
	(0.106)	(0.213)	(0.090)	(0.163)	(0.124)	(0.200)	(0.081)	(0.156)
Unionized	-1.236**	-1.513***	-0.985***	-0.766**	4.669***	4.754***	-2.091***	-2.150***
	(0.501)	(0.509)	(0.320)	(0.324)	(0.588)	(0.576)	(0.272)	(0.285)
Sector Equity Beta	0.069***	0.068***	-0.010	-0.009	0.014	0.008	-0.030***	-0.024***
	(0.021)	(0.023)	(0.009)	(0.008)	(0.009)	(0.011)	(0.003)	(0.004)
Share Retired	-0.288***	-0.250***	0.212***	0.138***	0.414***	0.402***	-0.276***	-0.233***
	(0.072)	(0.069)	(0.045)	(0.042)	(0.067)	(0.064)	(0.043)	(0.042)
ln(Avg. Acct. Bal.)	0.787***	0.710***	-0.165**	-0.075	0.423***	0.393***	-0.938***	-0.924***
	(0.089)	(0.089)	(0.072)	(0.071)	(0.082)	(0.080)	(0.061)	(0.062)
Observations	117,799	117,637	117,799	117,637	117,799	117,637	117,799	117,637
R^2	0.095	0.175	0.026	0.119	0.062	0.149	0.030	0.118
Year FE	X	X	X	X	X	X	X	X
NAICS FE		X		X		X		X
County FE		X		X		X		X

Notes: Table A5 displays the regression results corresponding to a linear regression model. The dependent variable is the portfolio weight of the corresponding asset class. The independent variables, other than the dummy variable *Union*, are all standardized such that they are in units of standard deviations. Observations are at the plan-by-year level over the period 2009-2019 where we restrict our attention to those smaller plans (below the median). Standard errors are clustered 2-digit NAICs by county level and are in parenthesis. *** p<0.01, ** p<0.05, * p<0.10.

Table A6: Relative Number of Equity Investment Option and Asset Allocation

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Relative No. Equity Options	19.4***	26.0***	26.6***	23.3***	28.3***	29.1***
	(0.74)	(0.74)	(0.78)	(0.78)	(0.77)	(0.79)
Offer Company Stock		5.63***	5.40***		5.77***	5.47***
		(0.16)	(0.18)		(0.16)	(0.17)
ln(Total Plan Asset)			0.12***			0.20***
			(0.045)			(0.048)
Observations	20,199	20,199	20,199	20,197	20,197	20,197
R^2	0.033	0.090	0.091	0.122	0.176	0.176
NAICS FE				X	X	X

Notes: Table A6 displays regression results of equity allocation on relative number of equity funds. Observations are at plan-by-year level over the period 2009-2019, weighted by total plan asset. We restrict plans whose start dates on 5500 Forms are on or after 2009. The dependent variable is equity allocation, which includes US equity, international equity and 50% of multi-asset funds. Relative No. of equity is computed following Benartzi and Thaler (2001), where each investment option is weighted by how long it has been in the plan and how well it has performed. To measure performance, we use S&P 500 Index as proxy for return on US equity, Barclays Agg Bond Index for bonds, S&P Global BMI for international equity, S&P US Treasury Bill 0-3 Month Index for cash/stable value. We assume return for multi-asset is 50% S&P 500 Index and 50% Barclays Agg Bond Index. For additional controls, we consider an indicator for whether the plan includes company stocks, log of total plan assets, and fixed effects for 2-digit NAICS code of sponsors of the plans. Standard errors are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A7: Alternative Model Specifications

(a) Risk Aversion and Expected Market Returns

	Obs	Mean	Std. Dev.	Median
Risk Aversion	243,268	3.553	0.472	3.506
Risk Aversion: No Time-Varying Intercept	243,268	3.599	0.286	3.604
Risk Aversion: Including Target Date Funds	243,268	3.554	0.482	3.515
Risk Aversion: New Plans Only	4,772	3.562	0.000	3.562
Risk Aversion: Simplified Risk Measure	243,268	7.630	1.685	7.160
Risk Aversion: 55 Factor Model	243,268	4.073	0.666	3.942
Expected Return	243,268	9.558	2.329	9.469
Expected Return: Time-Varying Intercept	243,268	9.696	2.150	9.766
Expected Return: Including Target Date Funds	243,268	9.729	2.219	9.615
Expected Return: New Plans Only	4,772	9.922	2.255	10.039
Expected Return: Simplified Risk Measure	243,268	14.951	4.068	14.445
Expected Return: 55 Factor Model	243,268	11.030	2.527	10.916

(b) Correlation: Baseline vs. Alternative Specifications

	Expected Return	Risk Aversion
Model: No Time-Varying Intercept	0.883***	0.576***
Model: Including Target Date Funds	0.948***	0.995***
Model: New Plans Only	0.844***	
Model: Simplified Risk Measure	0.700***	0.416***
Model: 55 Factor Model	0.817***	0.552***

Notes: Table A7 displays the results for our alternative model specifications. We estimate five alternative specifications. First, in the *No Time-Varying Intercept Model* we do allow mean risk aversion to vary year-by-year. Second, in the *Including Target Date Funds Model* we include target date funds when computing portfolios, risk, and expected market returns. Third, in the *New Plans Only Model* we estimate the model using data from 401(k)'s in the year the plan is introduced. We focus on the year of inception because it captures the active decisions of investors. Because we have fewer observations in this sample, we keep risk aversion constant across investors/plans in the *New Plans Only Model*. Fourth, in the *Simplified Risk Measure Model* we calculate the covariance of fund returns using a simplified factor model where we construct the factors by forming equal weighted portfolios based on the broad BrightScope categories reported in Table 1a. Fifth, in the *55 Factor Model* we calculate the covariance of fund returns using a 55 factor model where we construct our factors based on the Fama French 5 factors, 49 industry portfolios, and mommentum. Panel (a) displays mean, standard deviation, and median of the estimates of risk aversion and beliefs across our model specifications. Column (1) of Panel (b) displays the correlation between the estimated expected returns from the baseline model specification with the estimated expected returns from the other model specifications. Column (2) of Panel (b) displays the correlation between the estimated risk aversion from the baseline model specification with the estimated risk aversion from the other model specifications. Observations in both panels are at the plan-by-year level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A8: Financial Professionals - Expected Returns vs. Past Fund Returns

VARIABLES	(1)	(2)	(3)	(4)
Lag Fund Ret.	0.004***	0.016***	0.001	0.004***
	(0.001)	(0.003)	(0.008)	(0.001)
Lag Fund Ret. x New Investment				-0.000
				(0.001)
Observations	366,286	48,833	4,673	366,286
R^2	0.940	0.945	0.953	0.940
Multi-Level FE	X	X	X	X
New Funds		X		
New Plans			X	

Notes: Table A8 displays the regression results corresponding to a linear regression model. Observations are at the investment option-by-plan-by-year level over the period 2009 through 2019 where we restrict our attention to sponsors in the financial sector (NAICS 52). The dependent variable is the expected returns of the fund. Expected returns are calculated based on the specification reported in specification (3) of Table 5 where we set the dummy variable *Existing 401(k) Plan* equal to zero for each plan. Expected returns are winsorized at the 1% level. Each specification include plan-by-year, investment category (Morningstar$\times$BrightScope)-by-year, and index fund-by-year fixed effects. Standard errors are in parenthesis and are clustered at the plan level. *** p<0.01, ** p<0.05, * p<0.10.

Table A9: Financial Professionals - Predictability of Forecast Errors

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Lag Forecast Error	0.026***	-0.008**				
	(0.004)	(0.004)				
Lag Fund Ret.			-0.045***	-0.045***		
			(0.003)	(0.003)		
Change in Beliefs					-0.456***	-0.669***
					(0.026)	(0.040)
Observations	207,405	207,081	365,927	365,923	207,661	207,338
R^2	0.605	0.643	0.595	0.628	0.606	0.644
Year FE	X		X		X	
Plan×Year FE		X		X		X

Notes: Table A9 displays the regression results corresponding to a linear regression model (eq. 16). Observations are at the investment option-by-plan-by-year level over the period 2009 through 2019 where we restrict our attention to sponsors in the financial sector (NAICS 52). The dependent variable is investor's forecast errors as measured per eq. (15). *Lag Forecast Error* measures investors forecast error in the previous period. *Lag Fund Ret.* measures the annual fund return in the previous year. *Change in Beliefs* measures the change in investor's beliefs about the expected returns of the fund over the previous year. Standard errors are clustered at the plan level. *** p<0.01, ** p<0.05, * p<0.10.

Table A10: Portfolio Weights vs. Past Fund Returns

VARIABLES	(1)	(2)	(3)	(4)
Lag Fund Ret.	0.037***	0.029***	0.060***	0.045***
	(0.001)	(0.001)	(0.005)	(0.001)
Lag Fund Ret. x New Investment				-0.041***
				(0.000)
Observations	6,380,683	995,641	121,724	6,380,683
R^2	0.444	0.475	0.455	0.445
Multi-Level FE	X	X	X	X
New Funds		X		
New Plans			X	

Notes: Table A10 displays the regression results corresponding to a linear regression model. Observations are at the investment option-by-plan-by-year level over the period 2009 through 2019. The dependent variable is the weight of the fund in the investor's portfolio. Each specification include plan-by-year, investment category (Morningstar×BrightScope)-by-year, and index fund-by-year fixed effects. Standard errors are in parenthesis and are clustered at the plan level. *** p<0.01, ** p<0.05, * p<0.10.

Table A11: Stock Market Exposure vs. Local Economic Conditions

VARIABLES	(1)	(2)	(3)	(4)	(5)
Pop. Growth	-0.098**				-0.129**
	(0.047)				(0.051)
Home Price Growth		0.018*			0.014
		(0.009)			(0.011)
Establishment Growth			0.021		0.016
			(0.019)		(0.019)
GDP Growth				0.012*	0.013**
				(0.007)	(0.007)
Observations	407,714	425,206	431,589	424,859	394,986
R^2	0.767	0.763	0.762	0.762	0.767
Year FE	X	X	X	X	X
Plan FE	X	X	X	X	X

Notes: Table A11 displays the regression results corresponding to a linear regression model. Observations are at the investment plan-by-year level over the period 2009 through 2019. The dependent variable is the share in equities. Standard errors are in parenthesis and are clustered at the county-by-year level. *** p<0.01, ** p<0.05, * p<0.10.

Table A12: Stock Market Exposure vs. Employer Returns

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Firm Return (1 years)	0.004***	0.009***							0.005***	0.009***
	(0.001)	(0.002)							(0.001)	(0.002)
Firm Investment			-0.001	0.042**					0.007	0.038*
			(0.012)	(0.019)					(0.014)	(0.021)
Sales Growth					-0.000	0.012***			-0.001	0.005
					(0.002)	(0.004)			(0.003)	(0.004)
Employment Growth							0.003	0.011*	0.002	0.004
							(0.003)	(0.006)	(0.004)	(0.006)
Observations	20,155	20,315	17,931	18,067	19,755	19,891	19,711	19,852	17,403	17,521
R^2	0.803	0.131	0.777	0.134	0.805	0.130	0.805	0.129	0.781	0.133
Year FE	X		X		X		X		X	
Plan FE	X		X		X		X		X	
NAICS$\times$Year FE		X		X		X		X		X

Notes: Table A12 displays the regression results corresponding to a linear regression model. Observations are at the investment plan-by-year level over the period 2009 through 2019. The dependent variable is the share in equities. Standard errors are in parenthesis and are clustered at the plan level. *** p<0.01, ** p<0.05, * p<0.10.

Table A13: Return Predictability

VARIABLES	(1) Returns (1 yr)	(2) Returns (1-3 yr)	(3) Returns (1 yr)	(4) Returns (1-3 yr)
Beliefs	0.784***	2.607***	0.091	-0.281*
	(0.014)	(0.035)	(0.069)	(0.162)
Observations	79,172	68,854	79,172	68,854
R^2	0.458	0.413	0.684	0.674
Year FE	X	X	X	X
Risk Controls			X	X

Notes: Table A13 displays the regression results corresponding to a linear regression model (eq. 17). Observations are at the fund-by-year level over the period 2009 through 2019. The dependent variable is the future return measured over a 1 year and 3 year horizon and is annualized. Standard errors are in parenthesis and are clustered at the fund level. *** p<0.01, ** p<0.05, * p<0.10.

Table A14: Model Estimates Accounting for Labor Income Risk (θ and ψ)

VARIABLES	(1)	(2)
θ	-0.104***	
	(0.011)	
ψ	0.329***	0.096
	(0.022)	(0.078)
Observations	4,727,392	5,956,422
Plan×Year FE	X	X
Category×Year×Index FE	X	
Fund×Year FE		X

Notes: Table A14 displays the regression results corresponding to a linear regression model (eq. 19). Observations are at the investment option-by-plan-by-year level over the period 2009-2019. The dependent variable is the additional risk of investing a dollar in a given investment option, given the other portfolio holdings in the plan. We estimate each specification using 2-stage least squares. We instrument for expenses and the corresponding interaction terms using Hausman-type as described in the text. Because each observation reflects the average behavior of plan participants, we weight each observation by the total assets of the 401(k) plan. All specifications include plan-by-year, Morningstar investment category-by-BrightScope investment category-by-year fixed effects, and index-fund-by-year fixed effects. Standard errors are clustered at the plan level and are in parenthesis. *** p<0.01, ** p<0.05, * p<0.10.