FRANCISCO GOMES, MICHAEL HALIASSOS, TARUN RAMADORAI

Household Finance

Institute for Monetary and Financial Stability
GOETHE UNIVERSITY FRANKFURT

WORKING PAPER SERIES NO. 138 (2020)
Household Finance*

Francisco Gomes, Michael Haliassos, Tarun Ramadorai†

March 10, 2020

Abstract

Household financial decisions are complex, interdependent, and heterogeneous, and central to the functioning of the financial system. We present an overview of the rapidly expanding literature on household finance (with some important exceptions) and suggest directions for future research. We begin with the theory and empirics of asset market participation and asset allocation over the lifecycle. We then discuss household choices in insurance markets, trading behavior, decisions on retirement saving, and financial choices by retirees. We survey research on liabilities, including mortgage choice, refinancing, and default, and household behavior in unsecured credit markets, including credit cards and payday lending. We then connect the household to its social environment, including peer effects, cultural and hereditary factors, intra-household financial decision making, financial literacy, cognition and educational interventions. We also discuss literature on the provision and consumption of financial advice.

Keywords: Household finance

JEL Code: G50

*We gratefully acknowledge the helpful and constructive comments of the Editor and of four referees, as well as excellent research support by Tajda Spital and Johannes Wölfart.

†Gomes: London Business School and CEPR. Email gomes@london.edu. Haliassos: Goethe University Frankfurt, IMFS and CEPR. Email: haliassos@wiwi.uni-frankfurt.de. Ramadorai: Imperial Business School and CEPR. Email t.ramadorai@imperial.ac.uk.
1 Introduction

Households are central to our understanding of finance. They are the pivotal decision-making agents in financial markets, taking saving decisions which lead in turn to direct or indirect investments in financial markets, making them the ultimate owners of productive corporations in the economy. They also take borrowing decisions about loans that are sourced directly from other households or intermediated through financial institutions, and pay taxes to finance the activities of governments.

Despite the centrality of households to the financial system, for many years, the field of finance concerned itself almost exclusively with the study of financial markets, non-financial corporations, and financial institutions and other intermediaries. Households were often relegated to the background, modelled through a simplistic representative agent or as exogenous noise traders. In the broader field of economics, households received more attention, mainly focused on their demand for products, or on their savings behavior in a generic single asset.

A simplistic characterization of households is at variance with the hugely rich and complex reality. It also fails to take into the account the bewildering variety of options and choices available to individuals in modern financial markets. Such neglect of reality can significantly affect the predictions of theoretical models, and compromise the fit of empirical studies in both economics and finance. Carefully documenting and identifying the highly complex and heterogeneous nature of household financial decisions is also important for normative reasons. Such studies can help us to better understand the impact of different economic policies and/or regulations in a world with multiple constraints and imperfect markets. This is a central goal of the study of economics.

More recently, the importance of studying household decisions has been brought home sharply, as the international political landscape has been transformed by groups of individuals feeling disenfranchised for a number of reasons. One important factor underlying these recent developments is trends in income and wealth inequality. Household finance, by helping us to dig deeper into the causes and consequences of these trends, is therefore also relevant for the study of broader questions of political economy.
The image of household finance has changed significantly over the years, and more rapidly in recent times, owing to a combination of factors—demographic, historical, methodological, and technological. First, the importance of households in financial markets has been increasing, for example, as a result of the expansion of defined contribution pension plans which forces individuals to invest their retirement saving directly. In many cases, households are increasingly required to decide their own optimal retirement savings plans.

Second, partly as a consequence of events over the past few decades, society has developed a more profound appreciation of the importance of household financial decisions for both macro-economic outcomes and asset price determination. One such trend has been evident in the aftermath of the financial crisis, which has spurred significant and detailed analysis of households’ borrowing decisions in mortgage and other consumer credit markets.

A third important factor that has permitted more detailed analysis is the growing availability of large, detailed, and high-quality datasets documenting household financial behavior, and a concomitant explosion in techniques and computational power available to interrogate these datasets. This growth has not been limited to empirical work—technical advances and computational power have also allowed us to incorporate more realistic environments in our models and to take heterogeneity into account in a more serious fashion. Fourth, interest in more carefully modelling household preferences, beliefs, and constraints has also grown as a result of work in behavioral economics, which has forced a more serious questioning of prevailing orthodoxies in economic thought.

Finally, the digital revolution and the associated democratization and decentralization of finance potentially provide an unprecedented opportunity to address many shortcomings of household finance markets. However, these new developments also throw up associated challenges to both investors and suppliers as they grapple with new technology in an effort to optimize households’ lifetime trajectories using financial markets. These new developments have also provided a fillip to research in the field.

As a result of this rapid growth, the field of household finance is now so extensive that no single survey can do it justice.\footnote{We note, despite this growth in work in the area, that the institutional structure has been relatively slow}
comprehensive survey of all areas of household finance with a superficial treatment of each of them, or a more focused survey at the cost of not giving some very exciting topics the coverage that they deserve. We have chosen the latter approach; we caveat here that our choice of topics should not be viewed in any way as a reflection of their relative importance within the field. Rather, they reflect our own limitations—we focus primarily on those areas with which we are more familiar, and consequently have a comparative advantage at being able to describe to the interested reader. In the conclusion to this survey, we document what we believe are the principal open areas as a result of pursuing this strategy, and we explicitly acknowledge here that these and other important areas of household finance fully deserve to be the subject of future surveys by those more expert than ourselves.

The organization of the survey is as follows. Section 2 contains material on different saving motives for individuals and the allocation of those saving across different assets. This section includes an introduction to the basic infinite horizon consumption-saving problem and the lifecycle portfolio choice framework. It then discusses the literature on the observed low rates of asset market participation by households and the causes and consequences of this issue. This section also discusses evidence and theory on households’ asset allocation decisions conditional on participation, highlighting some new observations that have emerged from the data and in numerous studies. The section then briefly discusses household choices in insurance markets, surveys the literature on household trading behavior, and finally ends by surveying the literature on decisions on retirement saving and financial choices by retirees.

Section 3 moves from assets to liabilities, surveying the literature on mortgage choice, refinancing, and default, and discussing household behavior in unsecured credit markets, including credit cards and payday lending. Section 4 surveys the literature that connects the household to the social environment in which it operates, including work on peer effects. It then turns to discussing cultural and hereditary factors that influence household asset and liability decisions, and briefly surveys the literature on intra-household bargaining in

---

3

to catch up. For example, several major finance conferences do not yet have a separate household finance track for submissions, and the list of JEL codes has only recently been augmented with G5 for Household Finance.
household finance. Section 5 surveys the literature on financial literacy and cognition and their impacts on households, including the study of interventions in this space, Section 6 surveys the literature on the provision and consumption of financial advice, and Section 7 concludes.

2 Saving and Asset Allocation

In this section we discuss different saving motives for individuals and the allocation of those saving across different assets.

2.1 Saving motives

We first consider a simple frictionless infinite horizon consumption and saving problem for a power utility investor:

$$\max_{\{C_{t+s}\}} E_t \sum_{s=1}^{\infty} \beta^{s-1} \frac{(C_{t+s})^{1-\gamma}}{1-\gamma}$$

s.t. \quad W_{t+s+1} = R(W_{t+s} - C_{t+s}) + Y_{t+s}, \text{ and } W_s \text{ given}

where \(C_t\) and \(W_t\) denotes consumption and wealth, respectively; \(R\) is the (constant) rate of return on saving; and \(Y_t\) is a random variable capturing the agent’s endowment. The first order condition for this maximization problem yields the following Euler equation:

$$(C_t)^{-\gamma} = \beta R E_t(C_{t+1})^{-\gamma} \iff 1 = \beta R E_t \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma}$$

Taking a log-linear approximation and solving for expected consumption growth (with the notation \(g_t \equiv \ln(C_{t+1}/C_t)\), we obtain:

$$E_t(g_{t+1}) \simeq \frac{1}{\gamma} [\ln(R) + \ln(\beta)] + \frac{1}{2} \gamma \text{Var}[g_{t+1}]$$

Optimal saving is then determined by two terms. The first is the difference between the rate of return on saving and the discount rate, scaled by the elasticity of intertemporal

---

2 The parameters \(\gamma\) and \(\beta\) capture relative risk aversion and the subjective discount factor, respectively.

3 The equation holds exactly when consumption growth is log-normally distributed.
substitution, which determines optimal consumption smoothing. The second term captures precautionary saving and appears because we have uncertainty in the model. Zeldes (1989), Deaton (1991), and Aiyagari (1994) solve similar models while adding borrowing constraints, and where the endowment process, \( \{Y_s\}_{s=t}^{\infty} \), becomes labor income, which in Aiyagari (1994) is determined in equilibrium. Precautionary saving can arise as insurance, not just against labor income risk but also other sources of risk, for example medical expenditure risk as in Hubbard, Skinner, and Zeldes (1995), De Nardi, French, and Jones (2010), Ameriks, Caplin, Laufer, and Van Nieuwerburgh (2011), Yogo (2016) and Koijen, Van Nieuwerburgh and Yogo (2016).

Life-cycle models solving finite-horizon versions of the previous model are able to incorporate both retirement and a calibrated age-profile for labor income (e.g. Hubbard, Skinner, and Zeldes, 1995; Carroll, 1997; Gourinchas and Parker, 2002; Cagetti, 2003). More precisely, the labor income process is given by:

\[
Y_t = \begin{cases} 
  f(t) + \tilde{Y}_t & \text{for } t \leq K \\
  \bar{Y} & \text{for } t > K 
\end{cases}
\]

where \( K \) denotes the (exogenous) retirement age. Pre-retirement labor income is given by the combination of a deterministic function of age, \( f(t) \), and a random variable, \( \tilde{Y}_t \). After retirement, the household receives a constant income, \( \bar{Y} \). Throughout the paper we follow the common practice in the literature and refer to payments from both social security and defined benefit pensions as labor income during retirement, and therefore as part of human capital. Figure 1 plots a typical life-cycle profile with age-20 income normalized to 1. Average working life income ( \( E(Y_t) \) ) is typically much higher than retirement income ( \( \bar{Y} \) ), and in particular as \( t \) approaches \( K \). As a result, these models capture one additional saving motive for optimal consumption smoothing, namely, retirement saving.


\[\text{The discount rate is approximately } -\frac{\ln(\beta)}{\gamma}, \quad \text{and the elasticity of intertemporal substitution is } 1/\gamma.\]

\[\text{This particular profile is taken from Cocco, Gomes, and Maenhout (2005), and it was estimated using data from the Panel Study on Income Dynamics.}\]
determining wealth accumulation, particularly late in life. Introducing housing in the model discourages saving in financial assets, particularly in the presence of fixed costs of accessing those assets (Cocco, 2005). In models that also include a rental option, making home ownership an endogenous decision, there is an additional effect. Liquidity-constrained agents are forced to rent housing services, and therefore sufficient wealth accumulation is required to make an initial down-payment and to become a homeowner (Yao and Zhang, 2005). Finally, life-span uncertainty provides an incentive for households to decumulate wealth more slowly late in life, particularly in the presence of longevity risk (e.g. Horneff, Maurer, Mitchell, and Stamos, 2009; Cocco and Gomes, 2012; Yogo, 2016).

Saving rates in general, and retirement saving in particular, can also be negatively affected by introducing self-control problems in the model (e.g. Laibson, 1997; Angeletos, Laibson, Repetto, Tobacman, and Weinberg, 2001), and this can rationalize why, as discussed below, a significant fraction of the population does not save enough.

2.2 Baseline portfolio choice models

Before discussing the empirical evidence on asset allocation decisions, we introduce the results of different benchmark portfolio choice models, to help guide the discussion of both empirical findings and explanations for these findings.

2.2.1 Optimal portfolio choice without labor income

We first consider a simple static portfolio choice model with power utility preferences, two assets, and no frictions:

$$\max_{\alpha} E \left( \frac{(W_1)^{1-\gamma}}{1-\gamma} \right)$$

s.t.

$$W_1 = [\alpha R_1 + (1 - \alpha) R_f] W_0,$$

(1)

where $R_f$ is the return on a safe asset and $R_1$ is the return on a risk asset. Taking a log-linear approximation to the budget constraint, and assuming that wealth is lognormally

---

6Chetty and Szeidl (2007) extend this idea to the broader concept of consumption commitments.
distributed, the optimal portfolio share invested in the risky asset is given by:

\[
\alpha = \frac{E(r_1) + \frac{1}{2}\sigma^2_R - r_f}{\gamma \sigma^2_r},
\]

where \( r_1 \equiv \text{Ln}(R_1) \) and \( r_f \equiv \text{Ln}(R_f) \). The same expression was derived exactly in a continuous-time multi-period model with intermediate consumption by Merton (1969), and it is often referred to as the Merton rule.

### 2.2.2 Optimal portfolio choice with riskless labor income in a one-period model

For most households, human capital is the largest and most important asset. We therefore extend the simple portfolio choice model by replacing the budget constraint (1) with:

\[
W_1 = [\alpha R_1 + (1 - \alpha)R_f]W_0 + Y_1
\]

where \( Y_1 \) is labor income in period 1.

We first consider the case of riskless labor income, i.e., where \( Y_1 \) is a constant. The optimization problem is:

\[
Max \int \frac{([\alpha R_1 + (1 - \alpha)R_f]W_0 + Y_1)^{1-\gamma} f(R_1) dR_1}{1 - \gamma}
\]

where \( f \) denotes the pdf of the return distribution. The first order condition with respect to \( \alpha \) is:

\[
\int ([\alpha R_1 + (1 - \alpha)R_f]W_0 + Y_1)^{-\gamma} (R_1 - R_f)W_0 f(R_1) dR_1 = 0
\]

This can be re-written as:

\[
\int (\bar{\alpha} R_1 + (1 - \bar{\alpha})R_f)^{-\gamma} (R_1 - R_f) f(R_1) dR_1 = 0
\]

where

\[
\bar{\alpha} \equiv \frac{\alpha}{1 + \frac{Y_1}{R_fW_0}}
\]

The optimal \( \bar{\alpha} \) must be the same as the optimal \( \alpha \) with \( Y_1 = 0 \). So, denoting the latter by \( \alpha_{NY} \), we have:

\[
\frac{\alpha_Y}{1 + \frac{Y_1}{R_fW_0}} = \alpha_{NY} \iff \alpha_Y = \left(1 + \frac{PV(Y_1)}{W_0}\right) \alpha_{NY}
\]
where $\alpha_Y$ is the optimal solution with labor income.

From equation (2), we conclude that riskless (labor) income is like an extra endowment of the safe asset. Therefore the investor re-adjusts her financial portfolio to keep the share of total wealth invested in risky securities unchanged, relative to the one in the model without labor income (Merton, 1971).

### 2.2.3 Optimal portfolio choice with risky labor income

In reality future labor income is risky. Accounting for this brings two additional effects into the model; the optimal portfolio allocation is also determined by the correlation between future labor income realizations and stock returns, and the variance of labor income (see Heaton and Lucas, 1996; Viceira, 2001; Haliassos and Michaelides, 2003; Cocco, Gomes, and Maenhout, 2005).

**Correlation:** To understand the effect of correlation, we can consider the extreme case in which labor income is perfectly correlated with the return on the risky security. Intuitively, labor income is now equivalent to an endowment of the risky asset, and indeed if we repeated the algebra in the previous subsection, we would find that it is the share invested in the riskless asset that is scaled by one plus the ratio of human capital to (current) financial wealth.

Given the results from the cases in which there is either zero or perfect correlation, we can conclude that the net effect depends on the sign and magnitude of the correlation between stock returns and labor income shocks. This simple static model suggests the following rule of thumb: if the correlation between labor income and stock returns is lower (higher) than 0.5, then the portfolio share invested in stocks is higher (lower) in the presence of labor income.\(^7\) Campbell, Cocco, Gomes, and Maenhout (2001) and Davis and Willen (2013) find that this correlation is close to zero for the average household. Even taking into account heterogeneity across industries, the highest values of this correlation that they find still lie

---

\(^7\)This is just a rule of thumb since we also have the background risk effect discussed below, and because the stochastic process for labor income is not i.i.d.. In a multi-period model the correlation structure is often more complex.
Benzoni, Collin-Dufresne, and Goldstein (2007) argue that this correlation might be higher at lower frequencies, which is currently untestable given the state of time-series data availability on individual labor income. The empirical correlation between aggregate labor income and aggregate stock returns is quite high, but aggregate labor income fluctuations only account for about 10% of the total variation in individual-level labor income (see Campbell, Cocco, Gomes, and Maenhout, 2001). In addition, as discussed below, recent empirical evidence argues that there is no random walk component in individual labor income. A promising new approach is proposed by Catherine (2019), who documents a large correlation between higher-order moments of returns and income at business cycle frequencies, and shows that this has important implications for optimal portfolios.

**Background risk:** In classical portfolio choice theory idiosyncratic risk is irrelevant because agents hold diversified portfolios. However labor income risk is largely undiversifiable, i.e. agents cannot trade assets or write contracts to insure it. Under fairly general conditions (a utility function that exhibits decreasing and (weakly) convex absolute risk aversion), this gives rise to a background risk effect (Pratt and Zeckhauser, 1987; Kimball, 1993; Gollier and Pratt, 1996). The conclusion is that labor income risk, or any other source of background risk (e.g. expenditure shocks) crowds out asset allocation risk—Viceira (2001) extends this result to an infinite horizon portfolio choice model with intermediate consumption.

### 2.2.4 Baseline life-cycle portfolio choice model

We first present the basic life-cycle model of consumption and portfolio choice with un-spanned risky labor income and borrowing constraints introduced by Cocco, Gomes, and Maenhout (2005). The investor maximizes:

$$\max_{\{C_t, \alpha_t\}_{t=1}^T} \mathbb{E} \sum_{t=1}^T \left( \prod_{j=0}^{t-1} \pi_j \right) \beta^{t-1} \frac{C_t^{1-\gamma}}{1-\gamma}$$
s.t.

\[ W_{t+1} = \alpha_t R_{t+1} + (1 - \alpha_t) R_f (W_t - C_t) + Y_{t+1} \]

\[ W_t \geq 0, \quad \alpha_t \in [0, 1] \]

\[ R_t = \mu + \varepsilon_t, \quad \text{where} \quad \varepsilon_t \sim N(\mu, \sigma^2_R) \]

where \( \{ \pi_t \}_{t=1}^T \) are the conditional survival probabilities and the stochastic process for labor income before retirement (\( t < K \)) is given by:

\[
\ln(Y_t) = f(t, \theta) + P_t + U_t
\]

\[ P_t = P_{t-1} + Z_t, \quad Z_t \sim N(0, \sigma^2_Z) \]

\[ U_t \sim N(0, \sigma^2_U) \]

where \( f(t, \theta) \) is a function of age (\( t \)) and (potentially) other demographic characteristics (\( \theta \)).\(^8\) After retirement (\( t \geq K \)) labor income (i.e., payments from both social security and defined benefit pensions) is given by:

\[ \ln(Y_t) = \lambda[f(K) + P_K] \]

The model can be solved using backward induction with standard numerical techniques.\(^9\)

### 2.2.5 Consumption and portfolio allocations in the baseline life-cycle model

The previous model combines both retirement and precautionary saving motives. Figure 2 shows average simulated income, consumption and wealth accumulation for \( \gamma = 5, \beta = 0.97 \), and typical values for the other model parameters.\(^10\) Early in life, the household accumulates a small buffer stock to insure against negative income shocks. From around age 30-35 onwards, wealth accumulation accelerates rapidly, to finance consumption in retirement.

\(^8\)Gomes, Kotlikoff, and Viceira (2008) consider the case of endogenous labor supply before retirement.

\(^9\)Code available (in Matlab or Fortran) on http://faculty.london.edu/fgomes/research.html

\(^10\)For a range of reasonable values for the preference parameter of households see Calvet, Campbell, Gomes, and Sodini (2019). They structurally estimate a life-cycle model of consumption and portfolio choice with Epstein-Zin utility, using household-level data on both wealth accumulation and wealth composition.
After retirement, wealth converges to zero since there is no bequest motive in the model, and it does so quite rapidly since there are no medical expenditure shocks or longevity risk.

Figure 3 presents the optimal asset allocation before retirement, using the low correlation between labor income and stock returns in the baseline case in Cocco, Gomes, and Maenhout (2005). The portfolio rule is thus approximately given by equation (2) where $PV(Y_1)$ is the present-discounted value of all future labor income. As the investor ages, the crucial ratio $\frac{PV(Y_1)}{W_0}$ falls (since the numerator falls and the denominator rises) which leads to the decreasing age profile observed in Figure 3. However, Cocco, Gomes, and Maenhout (2005) discuss conditions under which the optimal equity allocation early in life can become an increasing function of $\frac{PV(Y_1)}{W_0}$ for low values of wealth. These conditions include high correlation between stock returns and labor income,\(^\text{11}\) or high levels of background risk.\(^\text{12}\) This pattern can also be obtained with significant negative skewness in returns (Fagereng, Gottlieb, and Guiso, 2017)\(^\text{13}\), or when investors have preferences which exhibit decreasing relative risk aversion, either through habit formation (Gomes and Michaelides, 2003; or Polkovnichenko, 2007) or non-homothetic preferences over basic and luxury consumption goods (Wachter and Yogo, 2010). Campanale, Fuggaza, and Gomes (2015) also obtain such a pattern by modelling differences in liquidity between stocks and bonds. Under any of those conditions the optimal equity exposure before retirement can become a hump-shaped function of age, since the increasing pattern of the policy function dominates early in life, when households have low wealth.\(^\text{14}\)

We now discuss the portfolio shares after retirement. We did not include this in Figure

\(^{11}\)Benzoni, Collin-Dufresne, and Goldstein (2007) consider that such high correlation might result from a low frequency association between the two variables, and thus it is not captured by the previous empirical studies.

\(^{12}\)Bagliano, Fugazza, and Nicodano (2017) and Catherine (2019) expand on this channel by considering persistent effects of negative income shocks, while using different specifications for the income process. Catherine (2019) combines this with negative skewness in stock returns and an (estimated) correlation between the two.

\(^{13}\)This large negative skewness is obtained by including idiosyncratic return shocks estimated from micro-data, and therefore does not require assuming (counter-factually) large negative market returns.

\(^{14}\)In some cases the increasing pattern can dominate all the way up to retirement.
2, for the simple reason that the model does not deliver unambiguous predictions here. As equation (2) shows, the equity allocation either increases or decreases with age, depending on what happens to the ratio $\frac{PV(Y_t)}{W_0}$. However, at retirement, both numerator and denominator are falling, so it depends on which decreases more rapidly. In the simple version of the model considered above, wealth falls quite rapidly, as shown in Figure 2, so the ratio $\frac{PV(Y_t)}{W_0}$ increases with age, leading to an increasing equity share. However, if we introduce bequest motives, medical expenditure risk, or longevity risk, this pattern can reverse. Finally, the preference parameters, and in particular the discount rate, are also very important for determining this particular result.

### 2.2.6 Further discussion of the income process

The income process that we described above is used in Carroll (1997), Carroll and Samwick (1997), Gourinchas and Parker (2002) and Cocco, Gomes, and Maenhout (2005), among others. As far as the parameters are concerned, Carroll and Samwick (1997) and Campbell, Cocco, Gomes, and Maenhout (2001) report estimates of the standard deviations of the transitory shocks (i.e., $\sigma_U$) for different education and occupation groups using data from the US Panel Study of Income Dynamics (PSID). These tend to decrease with education, and are typically between 0.05 and 0.35, although this might partially reflect measurement error in the survey data used for estimation. Estimates of the standard deviations of the permanent shocks (i.e., $\sigma_Z$) usually fall in the interval 0.05 to 0.2.

Some of these papers (e.g. Carroll and Samwick, 1997; Cocco, Gomes, and Maenhout, 2005) also consider a more general income process in which unemployment events are treated as separate states, occurring with some probability $p$:

$$\ln(Y_t) = \begin{cases} f(t, \theta) + P_t + U_t & \text{with prob. } 1 - p \\ Y & \text{with prob. } p \end{cases}$$

and where $Y$ is estimated to match labor income in a year with an unemployment spell. Others, like Hubbard, Skinner, and Zeldes (1995), for example, relax the assumption that
$P_t$ is a random walk, and instead consider:

$$P_t = \rho P_{t-1} + Z_t$$

This formulation adds one extra continuous state variable to the life-cycle model, since it is no longer possible to normalize by the current value of $P_t$; there is empirical evidence (e.g. Guvenen, 2007 and 2009) that favors this specification.

Other formulations include Lynch and Tan (2011), who consider time-variation in the mean and volatility of income growth and discuss how this influences the optimal equity exposure over the business cycle, and Meghir and Pistaferri (2004) and Storesletten, Telmer, and Yaron (2004), who document the importance of time-varying higher-order moments. Building on this, and using a large administrative dataset, Guvenen, Ozkan, and Song (2014) show that, while the variance of idiosyncratic shocks is acyclical, left-skewness increases in recessions. More precisely, they document that income dynamics are better characterized by a process where the idiosyncratic persistent shocks follows a mixture of normals, i.e.:

$$Z_t \sim \begin{cases} N(\mu_1, \sigma_1^2) \text{ with prob.} \pi \\ N(\mu_2, \sigma_2^2) \text{ with prob.} 1 - \pi \end{cases}$$

Finally, the model we have discussed so far assumes an exogenous labor supply and a fixed retirement date. Bodie, Merton, and Samuelson (1992) and Gomes, Kotlikoff, and Viceira (2008) relax the first assumption, while Fahri and Panageas (2007) allow for an endogenous retirement decision; Chai, Horneff, Maurer, and Mitchell (2011) relax both simultaneously. These additional margins of adjustments allow agents to improve consumption smoothing, but have a limited impact on their portfolio allocations.

### 2.3 Asset market participation

We now turn to surveying the literature on households’ observed participation in risky asset markets.
2.3.1 Equity market participation rates

The most important fact that emerges from analysis of the equity allocations of households is that a large fraction of the population simply does not own any stocks. This fact was documented by Mankiw and Zeldes (1991) and Haliassos and Bertaut (1995) for the U.S., while the articles in Guiso, Haliassos and Jappelli (2002) provide similar evidence for the U.K., Germany, Italy, and the Netherlands.

Guiso, Sapienza, and Zingales (2008), Christelis, Georgarakos, and Haliassos (2013), and Badarinza, Campbell, and Ramadorai (2016) document stock market participation rates in broader cross-sections of developed countries. Guiso, Sapienza, and Zingales (2008) use the 1997–1998 Financial Research Survey for U.K. data, the 2004 wave of the Survey for Health, Age, and Retirement (SHARE) for data on other European countries, and their U.S. data is taken from the 1998 Survey of Consumer Finances (SCF). Christelis, Georgarakos, and Haliassos (2013) take U.S. data from the Health and Retirement Survey (HRS), U.K. data from the 2004 English Longitudinal Study on Ageing (ELSA), and data from the 2004 and 2005 waves of SHARE for the other European countries. Finally, Badarinza, Campbell, and Ramadorai (2016) use data from the 2010 SCF for the U.S., data from the Eurosystem Household Finance and Consumption Survey (HFCS), Australian and Canadian data from their respective national household surveys, and employ national surveys for other European countries. In these papers, equity market participation rates are found to be below 50% for all countries except for Sweden (and Denmark in Christelis, Georgarakos, and Haliassos, 2013; though not in Guiso, Sapienza, and Zingales, 2008). The U.S. exhibits participation rates just below 50%, while the numbers for all other countries are far lower. The U.S. share has seemingly grown substantially from its 36.8% figure in 1983, but if households are removed that only invest in stocks through DC retirement accounts, the fraction participating in the U.S. in 2010 is closer to 20% (Badarinza, Campbell, and Ramadorai, 2016). Interest-

15 Typical estimates include assets held in defined contribution (DC) accounts. If we exclude these assets, then measured stock-market non-participation is substantially higher, as we discuss below.

16 The data for Australia, Canada, and selected European countries is for the year 2008, 2009, 2010, or 2012, depending on the specific country.
ingly this is essentially identical to the share seen in 1983 (see Haliassos and Bertaut, 1995).17 The seeming increase in stock market participation rates in the U.S. therefore appears largely driven by the growth of DC pensions.

2.3.2 Explanations for limited equity market participation

There has been voluminous literature on the factors contributing to limited stock market participation of households since the early work of Haliassos and Bertaut (1995), who had pointed to fixed costs and departures from expected utility. Based on this literature, explanations of limited stock market participation can broadly be divided into four categories, involving household preferences, costs (both pecuniary and informational) of participation, risks faced by households, and the influence of peers.

First, most common preference specifications (e.g., power utility) imply risk neutrality over infinitesimal gambles. This means that in the absence of other sources of risk or constraints, investors will always allocate a positive fraction of their wealth to an asset with a positive risk premium. However, non-participation can be obtained under preferences that exhibit first-order risk aversion such as disappointment aversion (Ang, Bekaert, and Liu, 2005), loss aversion (Gomes, 2005), narrow framing (Barberis, Huang, and Thaler, 2006), rank-dependent utility (Chapman and Polkovnichenko, 2009), ambiguity aversion (Cao, Wang, and Zhang, 2005; Campanale, 2011; Peijnenburg, 2018), or news-utility (Pagel, 2018).

Second, zero stockholdings can also arise if there are fixed costs of participation. As Vissing-Jorgensen (2002), Haliassos and Michaelides (2003), and Gomes and Michaelides (2005) show, since a significant fraction of the population accumulates very little investable wealth, even a small participation cost is enough to deter households from investing in stocks. These papers are able to match empirically observed participation rates using costs such as explicit monetary expenditures (e.g. setting up a brokerage account), as well as indirect costs. These indirect costs include explicit and implicit costs of learning about the stock

\[17\] The percentage of the U.S. population who invest in stocks outside of retirement accounts increased to 32% in the 2001 SCF sample, but it subsequently declined again, as these numbers show.
market or about how to invest, costs of setting up an investment account, and/or the costs of finding an investment advisor or seeking advice—these indirect costs can be especially high for individuals with low financial literacy.

Third, turning to risk-based explanations, zero equity allocations can be obtained in models in which labor income is highly correlated with stocks (e.g. Benzoni, Collin-Dufresne, and Goldstein, 2007). Guiso, Sapienza, and Zingales (2008) highlight the importance of lack of trust in the stock market and/or financial institutions which can limit the expected return via a perceived probability to be cheated, and increase the perceived risk of stock market investments, decreasing participation rates.

Finally, individuals might also be more willing to participate in the stock market if their peers are also stock market participants, due either to social utility, or the imitation channel (Hong, Kubik, and Stein, 2004), as discussed more extensively in section 4.1 on peer effects. Empirical evidence provides support for the effects of all of these channels on stock market participation decisions, and we turn to describing this evidence next. As in most areas of household finance, future research is strongly encouraged to explore and explain the multiple sources of heterogeneity observed in the data.

### 2.3.3 Empirical evidence for the different mechanisms

In support of preference-based explanations, Dimmock, Kouwenberg, Mitchell, and Peijnenburg (2016) show that measures of ambiguity aversion are negatively related to stock market participation. Calvet, Celerier, Sodini, and Valee (2019) show that the broad demand for capital protected investments in Sweden can be explained in a model where investors have first-order risk aversion with narrow framing (Barberis and Huang, 2009).

In support of the information cost channel, there is evidence that individuals are more likely to participate in the stock market when they have high financial literacy (Calvet, Campbell, and Sodini, 2007; and van Rooij, Lusardi, and Alessie, 2011), higher education (Black, Devereux, Lundborg, and Majlesi, 2018), or high IQ (Grinblatt, Keloharju, and Linnainmaa, 2011).\(^\text{18}\) Calvet, Campbell, and Sodini (2007) also show that non-participants

\(^{18}\)Both set of results are obtained after controlling for education, wealth, and income.
would likely be inefficient investors, and as a result, a small cost of participation would be enough to deter them from such investments. Households are also more likely to invest in stocks when they are better informed about financial markets (Guiso, Haliassos, and Jappelli, 2002) or when their neighbors have experienced good stock returns (Kaustia and Knüpfer, 2012)—this result can also be interpreted as evidence in favor of peer effects. We defer discussion of empirical research on peer effects to the relevant section 4.1 of this survey.

Participation cost-based explanations receive strong support from the very high observed correlation between participation rates and wealth (see, for example, Guiso, Haliassos, and Jappelli, 2002; Campbell, 2006; and Guiso and Sodini, 2013). While other channels could also be responsible for the positive correlation between these variables, the participation costs theory produces this result most strongly and clearly, since poor households will not find it worthwhile to incur fixed costs, thus optimally choosing to not invest in stocks. However, while participation costs can explain non-participation for the poor, they cannot explain why a non-trivial fraction of wealthy households also does not invest in stocks. Explanations based on trust, non-standard preferences, or peer effects can plug this gap. Such explanations can also help to rationalize the findings in Andersen and Nielsen (2011) and Briggs, Cesarini, Lindqvist, and Östling (2015), who study households who receive large inheritances and large lottery gains, respectively. They show that for households who were previously non-stockholders, less than half experiencing such random increases in wealth subsequently decide to invest in stocks.

Bonaparte, Korniotis, and Kumar (2018) document that about 25% of stockholders in the U.S. enter/exit non-retirement investment accounts biennially. They replicate these entry and exit patterns in a life-cycle model with per-period participation costs. Risk-based explanations could only deliver time-variation in participation if they also consider for time-variation in (expected) risk or in the price of risk. For example, time-varying first-order risk-aversion, or constant first-order risk aversion combined with time-varying beliefs about future expected returns.

Choi and Robertson (2020) survey households on the factors driving their asset allocation decisions. Among those that did not invest in stocks, the main explanation provided for their
decision was “The amount of money I have available to invest in stocks is too small” (46.2% of the sample) followed by “I don’t like to think about my finances” (35.6% of the sample). Both of these support the (monetary and information) cost channels. They also report that trust concerns are important factors; and Guiso, Sapienza, and Zingales (2008) and Giannetti and Wang (2016) also show that trust levels can explain differences in participation rates in the data.19

As discussed earlier, stock market participation rates have increased significantly over time, particularly in the U.S., where this tendency has almost exclusively been driven by the expansion of DC pension plans. This provides strong evidence in favor of the participation cost channel. DC plans have significantly decreased the (direct and indirect) cost of stock market participation, particularly for households with limited financial literacy. The peer effects and trust channels are also potentially contributing factors. It is also plausible that DC plans are perceived as more trustworthy investment vehicles than direct or delegated direct participation. On the other hand preference-based explanations, or risk-based explanations would struggle to explain this large shift in participation rates. It is hard to argue that these plans have (significantly) changed the preferences of investors, or the correlation of their labor income with stock returns.

2.3.4 Other factors influencing equity market participation

In theory, stock market participation should also be influenced by factors that determine the optimal equity allocation conditional on participation, even if those factors cannot deliver non-participation in isolation. In the absence of the frictions or factors considered earlier, theoretically, investors should participate in the stock market given a positive expected equity premium. However, if some or all of the factors inhibiting participation are present, declines in the expected equity premium should in theory make households less likely to invest in stocks, as the introduction of some expected cost of participation makes the net benefit of participation less attractive. Consistent with this intuition, Malmendier and Nagel (2011)

---

19Choi and Robertson (2020) also include questions about social norms and peer effects but those factors are hard to capture, and individuals might be less likely to admit to them in a survey.
and Hurd, Van Rooij, and Winter (2011) find a positive correlation between expected stock market returns and stock market participation, while Bonaparte, Korniotis, and Kumar (2014) show that individuals are more likely to participate in the stock market when they face less income risk or when their income growth is less correlated with stock returns.

Of course, agents’ expectations of stock returns could also vary based on their beliefs. In support of this, Malmendier and Nagel (2011) show evidence that household-level differences in expected returns are influenced by past experiences (i.e., stock market returns experienced by investors over their lives). However, Knüpfer, Rantapuska, and Sarvimäki (2017a) document that negative labor market shocks have a permanent negative impact on stock market participation, suggesting that the effect of negative experiences on participation might arise from more than just reductions in their future expectations of returns.

Households might also find it optimal not to invest in stocks if they have large amounts of debt, either from mortgages (Cocco, 2005; and Yao and Zhang, 2005), or from high-interest uncollateralized debt (Davis, Kubler and Willen, 2006), since debt can reduce their net investable wealth. Without additional features, these frameworks can only explain non-participation by households with zero wealth. However, adding in small participation costs to these models predicts low financial wealth accumulation by a significant fraction of the population, meaning that they can rationalize zero equity investments for a larger fraction of the population (Cocco, 2005).

Ameriks and Zeldes (2004) and Poterba and Samwick (1997) estimate significant age effects in participation rates, but the exact patterns of stock market participation over the life-cycle are hard to pin down, as we discuss in more detail below. Depending on alternative identification assumptions, Ameriks and Zeldes (2004) find evidence for either an increasing pattern almost until retirement and flat thereafter, or flat almost until retirement and decreasing thereafter. In contrast, Poterba and Samwick (1997) and Fagereng, Gottlieb, and Guiso (2017) estimate hump-shaped patterns, peaking around age 40 or age 60, respectively.
2.3.5 Participation in other assets and the “asset tangibility preference”

Guiso, Haliassos, and Jappelli (2002), Christelis, Georgarakos, and Haliassos (2013) and Badarinza, Campbell, and Ramadorai (2016) report participation rates for other asset classes across households in developed countries. The percentage of households owning safe assets (such as bank deposits) is above 90% in almost all countries, and the percentage of households investing in bonds is typically below 10% with the exception of Italy (14.6%), the UK (28.2%), and the U.S. (13.2%).

Ownership in real estate is large, and varies significantly across countries, from as low as 44.2% (Germany) to 89.9% (Slovakia). Finally, roughly 10% of households have some form of private business—the country with the highest value reported is Italy, with 18.4% (Badarinza, Campbell, and Ramadorai, 2016). Interestingly, Christelis, Georgarakos, and Haliassos (2013) document that in Europe, countries with the highest homeownership rates, somewhat counterintuitively, tend to have the lowest mortgage participation rates. The picture of participation differs substantially for the sample of developing economies (accounting for over 50% of the world’s population) considered by Badarinza, Balasubramaniam, and Ramadorai (2019). On average, only 60% of households in these countries hold any form of financial asset, while real estate ownership is above 75% in every country considered.

The substantial share of real (or “tangible”) assets and the correspondingly low share of financial assets in a number of countries cutting across the development landscape can be thought of as revealing a “household asset tangibility preference”. Perhaps one explanation for this result is that the value of tangible assets tends to be greater in environments with low financial development, since collateral is potentially more valuable when screening borrowers is costly. This aspect is worth further exploration, but it is unlikely to be the only one, as preference for “tangibility” is also present in some developed countries, such as Greece and Italy. A complementary explanation may be that intergenerational transfers are more important in environments in which retirement and other saving markets are not efficient, public health and retirement systems are not reliable, and housing provides an easy way

---

20 For models with optimal bond holdings see Campbell and Viceira (2001), Van Hemert (2009), Kojien, Nijman, and Werker (2010) and Munk and Sorensen (2010).
to transfer wealth across time and generations. We believe further investigation of this empirical evidence is warranted.

Finally, an important asset class for households that has received comparatively less attention, even in developed countries, is insurance products, namely life-insurance, home insurance, and insurance for other durable goods. In addition, in countries where the level of coverage and/or quality of medical service financed by the government is not very high, health and medical insurance are also very important. In those countries, the cost of out-of-pocket medical expenditures, which includes the cost of long-term health maintenance or nursing care in old age can be very large and has been increasing over time, (see, for example, De Nardi, French, and Jones, 2010; Brown and Finkelstein, 2011; Ameriks, Caplin, Laufer, and Van Nieuwerburgh, 2011).

Most empirical studies document that individuals’ life insurance holdings tend to increase with income (e.g. Truett and Truett, 1990). Bernheim (1991) and Chen, Wong, and Lee (2001) find that the demand for life insurance tends to decrease with age, but Truett and Truett (1990) obtain the opposite result. Inkmann and Michaelides (2012) find that the demand for term insurance is positively correlated with measures of bequest motives, and Truett and Truett (1990) and Bernheim (1991) find that individuals with more children and higher education, respectively, have a higher demand for life insurance products. There is more limited evidence on how other forms of insurance relate to the economic and/or demographic characteristics of individuals. One exception is Hugonnier, Pelgrin, and St-Amour (2013), who find that the demand for health insurance increases with wealth. For theoretical models of household demand for insurance see, for example, Bernheim (1991), Brown, and Finkelstein (2008), Davidoff (2010), Inkmann and Michaelides (2012), Hugonnier, Pelgrin, and St-Amour (2013), Hubener, Maurer, and Rogalla (2014), Koijen, Van Nieuwerburgh, and Yogo (2016), and Cole, Kim, and Kruegger (2019).\footnote{Yogo (2016) studies investments in health status.}
2.4 Asset allocation decisions of asset holders

In this subsection we discuss the asset allocation decisions of individuals conditional on participation in asset markets. We then move to discussing the trading behavior of investors, primarily in equity markets, following which, we discuss household investments in retirement accounts.

2.4.1 Average portfolio shares

Guiso, Haliassos, and Jappelli (2002) report the share of financial wealth invested in risky assets by stock market participants for a sample of four countries: the U.S., Netherlands, Germany and Italy. When risky assets are defined as total equity holdings, this conditional “risky share” is 59.6%, 53.6%, 21.8%, and 57.3% respectively for each of those countries.22

Badarinza, Campbell, and Ramadorai (2016) consider shares of total wealth for a cross-section of 13 countries. The largest asset class is real estate, averaging 55.1% of total wealth. There is, however, significant cross-country heterogeneity, with the minimum of 36.6% (Germany) and the maximum of 77.4% (Slovakia), once again showing evidence that tangible asset holdings appear related to the degree of a country’s development. Financial assets constitute on average 27% of total household wealth, but again there is substantial heterogeneity, with a minimum of 10.0% (Slovenia) and a maximum of 47.9% (Germany). Directly held stocks, mutual funds, and bonds correspond to only about 0.9%, 0.9%, and 0.4% of total household wealth, with maximum values of 2.8%, 2.4%, and 1.6% respectively. These numbers largely reflect the previously documented limited participation in those asset classes, as well as the asset tangibility preference.

2.4.2 Life-cycle effects

In the household survey of Choi and Robertson (2020), “years left until retirement” is the most cited factor as a determinant of equity shares by stock market participants, so life-

22They also have information for the UK, where the conditional share is 42.7%, but here their data do not include holdings through mutual funds and retirement accounts.
cycle considerations clearly appear important to households. As discussed in section 2.2.3, the simple life-cycle asset allocation model implies that the risky share of financial wealth should be a decreasing function of age until retirement, with no clear predictions about behavior after retirement. However, as we also discussed earlier, several extensions of the model can generate a hump-shaped pattern or even a monotonically rising one (e.g. Cocco, Gomes, and Maenhout, 2005; Benzoni, Collin-Dufresne, and Goldstein, 2007; Campanale, Fuggaza, and Gomes, 2015; Fagereng, Gottlieb, and Guiso, 2017; Bagliano, Fugazza, and Nicodano, 2017; Catherine, 2019).

Estimating the life-cycle profile of the risky share in financial assets faces the classic identification problem of disentangling age, time, and cohort effects—the relationship “current year” = “birth year” + “age” implies perfect collinearity between the three variables, making it impossible to identify all of them simultaneously. Researchers have exploited three different approaches to address this problem. One alternative is to impose specific parametric restrictions on one (or more) of the three effects. Another option is to model one or more of these effects using specific variables which are not collinear with the others. The third alternative is to assume that the relationship is non-linear.

Poterba and Samwick (1997) find strong evidence for cohort effects in the data. They do not estimate risky shares conditional on participation, but find very similar hump-shaped patterns for both the unconditional equity share as well as stock market participation rates, with both peaking at around age 40. Their results suggest a relatively flat conditional risky share over the life cycle. Ameriks and Zeldes (2004) attempt to identify age effects by considering two special cases, involving either no cohort effects or no time effects. They find that the results under those two assumptions are very different—with cohort effects, the risky share of stock market participants increases with age until around age 60 and then flattens; and with time effects, the equity share appears flat until around age 55 and decreases subsequently.

More recent evidence is available in Fagereng, Gottlieb, and Guiso (2017), who use Norwegian administrative data which they argue is subject to less measurement error. They 23An extreme example is to completely shut down one of them by setting it equal to zero.
address the identification problem using two alternative formulations. In the first approach they model cohort effects based on individuals’ life-time experiences with returns (motivated by the results in Malmendier and Nagel, 2011), and in the second approach, they impose the restriction that time effects sum to zero once the variables have been detrended (following Deaton and Paxson, 1994). Using both approaches, they find that the risky asset share of stock market participants is a decreasing function of age. However, in contrast, using a similar approach, and U.S. SCF data, Catherine (2019) obtains a flat or mildly increasing pattern in the risky share with age. These contrasting results could reflect differences between the U.S.A and Norway, but they might also reflect measurement error in the self-reported holdings recorded in the SCF, as opposed to the administrative data employed in Fagereng, Gottlieb, and Guiso (2017).

2.4.3 Income risk and borrowing constraints

Consistent with the implications of the life-cycle models, Guiso, Jappelli, and Terlizzese (1996) find that both income risk and borrowing constraints significantly decrease risky asset holdings, while Bonaparte, Korniotis, and Kumar (2014) find that risky shares are lower for households with more income risk, or for household where income growth is more correlated with stock returns. Using identification from twins to control for latent forms of heterogeneity, Calvet and Sodini (2014) also conclude that income risk decreases risky asset shares. On the other hand, Massa and Simonov (2006) find that most investors tend to hold stocks that are positively correlated with their labor income, although they find that this pattern is weaker or even reversed for wealthy investors. Further evidence is provided by Choi and Robertson (2020), where borrowing, labor income, and other risks are listed among the factors determining equity allocation decisions. Fagereng, Guiso, and Pistaferri (2018) exploit the insight that the estimated impact of background risk on portfolio choice depends on the quantity of risk as well as on investors’ sensitivity to such risk. They identify these two effects separately in the data, and find that risky shares are, in fact, very

---

24 In the same way that the risk premium on a security depends on the price of risk and the quantity of risk.
sensitive to changes in background risk. The total estimated effect is small primarily because
the quantity of this risk faced by investors is small on average.

2.4.4 Housing

If we exclude the present-value of future labor income, housing is by far the most important
asset for most households (see section 2.4.1). Glaeser and Gyourko (2018) provide a useful
survey of the effects of housing prices on household wealth, Piazzesi and Schneider (2016)
discuss housing in macroeconomics, and Goodman and Mayer (2018) take a detailed look at
homeownership in the U.S. Going back to the asset tangibility preference, it is also important
to connect housing to risky asset shares. In this vein, Cocco (2005) and Yao and Zhang
(2005) solve life-cycle models with both portfolio choice and housing decisions. They
show that, controlling for wealth, individuals for whom real estate is a higher fraction of
their total wealth should invest less in risky assets. In the presence of adjustment costs,
housing represents a consumption commitment, and as such increases effective risk aversion
(Grossman and Laroque, 1990; Chetty and Szeidl, 2007). In contrast, those with large
mortgage balances (again relative to their total wealth) should have a higher risky asset
share. Cocco (2005) and Yao and Zhang (2005) show that those predictions are confirmed in
the data, a result also obtained in Heaton and Lucas (2000). Additional evidence for these
channels is provided in Calvet and Sodini (2014) and Chetty, Sandor, and Szeidl (2017).

2.4.5 Human capital and wealth

As discussed in section 2.2.3 life-cycle models predict that human capital is a close substitute
for bonds rather than stocks, unless the correlation between labor income shocks and stock

---

25See also Fischer and Stamos (2013) and Corradin, Fillat, and Vergara-Alert (2014).
26Sinai and Souleles (2005) argue that home ownership provides a hedge against rental fluctuations, and
that price risk is limited by the geographical correlation of housing prices.
27Numerous articles tackle the effects of homeownership on household consumption and the factors de-
dtermining house purchase and sale decisions—a partial sample includes Mian, Rao, and Sufi (2013), Bailey,
Cao, Kuchler, and Stroebel (2018), Benmelech, Guren, and Meltzer (2019), Andersen, Badarinza, Liu, Marx,
and Ramadorai (2019), and references therein.
returns is very high. Therefore, for a given value of human capital, the standard model implies that the risky share should be a decreasing function of wealth. However, as also mentioned in section 2.2.3., several extensions of the model can deliver an increasing pattern, at least over certain levels of wealth.

In the data, wealth and human capital are highly correlated, which makes it difficult to disentangle the effect of these two variables separately. The Swedish registry data used in Calvet and Sodini (2014) reports both asset holdings and income, allowing them to construct accurate measures of human capital at the individual level. They find that a higher present-value of human capital increases risk-taking, consistent with the conclusion that human capital is a closer substitute for bonds than stocks.

As far as wealth is concerned, the evidence is more mixed. Calvet and Sodini (2014) estimate a risky financial share that is an increasing and concave function of financial wealth, while Wachter and Yogo (2010) and Heaton and Lucas (2000), using U.S. SCF data and tax return data, also find that the risky share increases with wealth. However, using U.S. PSID data, Brunnermeier and Nagel (2008) find that the risky share in liquid wealth is flat in wealth, a result also obtained by Chiappori and Paiella (2011) using data from the Bank of Italy’s Survey of Household Income and Wealth.

Campbell (2006), Bach, Calvet, and Sodini (2019), and Fagereng, Guiso, Malacrino, and Pistaferri (2020) report wealth shares by asset class for different percentiles of the wealth distribution for the U.S., Sweden, and Norway, respectively. Safe assets account for the majority of the wealth of poorer households, but higher in the wealth distribution, housing becomes the dominant asset. At the right tail of the distribution, private equity becomes very important, and it surpasses housing for those in the top 1% (see also Heaton and Lucas, 2000). The share of public equity in financial wealth is particularly low for the poor, mainly because of low participation rates. This is also true for the richest 1%, who have a large share of their wealth in private equity. Excluding those two groups, the public equity share is mildly increasing in wealth.

---

28 The data in Campbell (2006) also includes vehicles, a very important asset for those at the left tail of the wealth distribution.
2.4.6 Other factors

We conclude this section by briefly summarizing evidence on other factors that have also been shown to have a significant impact on risky shares. Guiso, Jappelli, and Terlizzese (1996), Guiso, Sapienza, and Zingales (2008), Calvet and Sodini (2014), and Dimmock, Kouwenberg, Mitchell, and Peijnenburg (2016) respectively document that health risk, trust levels, internal habits and ambiguity aversion all influence equity holdings. These factors are also present in the survey of Choi and Robertson (2020), along with other factors such as “return disasters” (consistent with the models of Fagereng, Gottlieb, and Guiso, 2017; Catherine, 2019), and return expectations (in line with the evidence of Malmendier and Nagel, 2011). Black, Devereux, Lundborg, and Majlesi (2018) connect levels of education with risky shares, showing that more educated households invest more in stocks.

With regards to other demographic factors, Campbell (2006) and Calvet and Sodini (2014) show that household size is associated with a lower risky share, and Love (2010) and Hubener, Maurer, and Mitchell (2016) provide model-motivated evidence on the impact of changes in family size and marital status transitions. Relatedly, Heimer, Myrseth, and Schoenle (2019) and Balasubramaniam (2018) show that mortality expectations have an important impact on saving behavior and portfolio allocations.

2.4.7 Under-diversification

Using U.S. income tax data, Blume and Friend (1975) document that most investors hold highly undiversified portfolios, contrary to the implications of standard rational models of portfolio choice. Subsequent evidence connects this fact to familiarity bias, i.e., the hypothesis that households tend to buy stocks of companies with which they are more closely acquainted. For example, French and Poterba (1991) and Grinblatt and Keloharju (2001a) document that individual investors are more likely to hold stocks of companies based in their own countries. Even within the same country, investors appear to over-invest in local companies (Huberman, 2001; Feng and Seasholes, 2004; Ivković and Weisbenner, 2005).

\[29\] As discussed below, this was before the proliferation of index investing.
Knüpfers, Keloharju, and Linnainmaa (2012) find that investors are more likely to buy shares of companies whose products and services they consume. Other findings show that investors tend to allocate a significant fraction of their equity holdings to their own employer’s stocks, and that this tendency cannot fully be explained by constraints on their holdings (Benartzi, 2001; Mitchell and Utkus, 2003; Poterba, 2003).

The proliferation of relatively cheap index tracker vehicles has seemingly also contributed to an increase in diversification over time. Using comprehensive data on all financial asset holdings of Swedish households, Calvet, Campbell, and Sodini (2007) show that while households do appear to hold undiversified positions in a few individual stocks, either this behavior is concentrated in households with small risky asset shares (the extensive margin is small), or this only represents a small fraction of holdings for households who mainly invest in well-diversified mutual funds (the intensive margin is also small). They conclude that most investors therefore suffer very small implied return losses from under-diversification.

Rational under-diversification can be obtained under rank-dependent preferences (see, e.g., Polkovnichenko 2005), preference for skewness (Mitton and Vorkink, 2007), anticipatory utility (Brunnermeier, Gollier, and Parker, 2007), preferences with social status concerns (Roussanov, 2010), or ambiguity aversion (Dimmock, Kouwenberg, Mitchell, and Peijnenburg, 2016). It can also arise under probability weighting prospect theory (Barberis and Huang, 2008). However, preference-based explanations find it hard to match the strong evidence in favor of a familiarity bias. Van Nieuwerburgh and Veldkamp (2010) argue that this could be the result of investors tilting their portfolios towards stocks about which they have an informational advantage. Knüpfers, Rantapuska, and Sarvimäki (2017b) find that stockholders tend to hold the same securities as do their parents. Furthermore, this effect is stronger when they are more likely to communicate with each other, and when they are more susceptible to social influence, thus providing support for both the information channel and peer effects, respectively. Bekaert, Hoyem, Hu, and Ravina (2017) show using a large dataset of individual choices in 401(k) plans that education, financial literacy, and exposure to foreign-born individuals are associated with greater international diversification, also suggesting that familiarity and information-based explanations may hold part of the answer.
This section has highlighted the wide range of factors that have been shown to influence asset allocation decisions of households. The increasing availability of detailed and comprehensive data which combines information on individual portfolios and multiple other individual characteristics, makes this very promising research area going forward. This literature also emphasizes the importance, on the theory side, of developing models that not only try to incorporate all of these important factors, but also models that take household heterogeneity seriously.

We now turn to surveying the large literature on household portfolio construction and maintenance, and the evidence on household trading behavior in the stock market.

2.5 Trading

As discussed earlier, stock market participation is clearly beneficial to many households, at least in theory. Yet, in practice, gains from participation can depend on the precise manner in which households participate—with deviations from idealized participation potentially imposing costs that can reduce or even outweigh accrued benefits. One particularly tricky area for households is portfolio construction and maintenance, which can involve households trading in and out of stocks.

To take just one example, while Markowitz (1952) outlines the benefits of portfolio diversification, even this maxim is anything but simple when it comes to the details of implementation. Immediate questions arise even if the household’s goal is to trade in such a way as to achieve diversification: For example, how frequently should investors rebalance their portfolios, especially in the presence of transactions costs to achieve this goal (e.g., Dumas and Luciano, 1991)? How should households trade in the presence of specific constraints that need to be factored into their portfolio weights, such as short sales constraints (see, e.g., Jagannathan and Ma, 2003)? What trading frequency should households operate at, assuming they must estimate means and especially variances of asset returns, especially if expected returns are time-varying (e.g. Barberis, 2000; Ledoit and Wolf, 2003)?

Many additional complications exist, making the details of stock investing and trading a
veritable minefield for households. First, while we have earlier discussed the availability of low-cost index tracker funds, this is not widespread. In many economies, low-cost diversification vehicles such as passive index funds or their more sophisticated equivalents simply do not exist, and so households must invest directly in stocks (Badarinza, Balasabramaniam, and Ramadorai, 2019; Campbell, Ramadorai, and Ranish, 2019). This means that in many situations, households must trade to achieve diversified portfolios. Second, fund selection, even in a market in which low-cost diversification vehicles do exist can be just as complex and confusing for households given that the number of funds is greater than the number of individual stocks in many markets around the world, and details of fees and performance can be difficult to compare across vehicles (Anagol and Kim, 2012). So trading or at least fund selection appears inevitable even when managed vehicles are available. Third, high-quality financial advice or even delegation of these tasks is often only accessible to households capable of paying significant fees to access such services. Perhaps more worryingly, much recent work, reviewed in the section on financial advice below, has demonstrated significant conflicts of interest, sometimes even manifested in adviser misconduct, that plague the provision of such advice.

Fourth, transactions costs can be substantial, especially in illiquid stocks, which introduces a significant wedge between promised and actual gains from stock market, and makes optimal trading behavior hugely complex. Finally, households face considerable cognitive challenges when confronted with a task involving a vast amount of information and complicated decisions, making cognitive ability a relevant factor for stock trading (see below).

As might be expected, the pursuit of explanations for household trading behavior that focus on cognitive limitations has generated considerable overlap between the literature on household trading behavior and that in behavioral finance and economics (see Barberis, 2018, for an excellent survey). This growing literature offers a virtually unparalleled level of detail about household decisions—reconstructed from increasingly precise trading records sourced from around the world, and as such, it allows for “in-the-field” validation of the vast and

---

30 There are now more indexes than stocks, Bloomberg, 2017 (shorturl.at/mAOR1)
growing lab experimental literature on household decision-making.\footnote{It is worth noting here that this literature studies the determinants and performance of equity trades, rather than equity portfolio composition, using higher-frequency administrative data that often do not reveal the composition of total household portfolios.}

In some cases, investors are prone to hyperactivity, trading substantially more than can be justified by their expected returns, and in others, they are prone to torpor, failing to realize losses or rebalancing portfolios far less frequently than predicted by optimizing models. The search for the underlying causes of these seemingly different phenomena have generated significant theoretical work in addition to careful empirical documentation of behavior. We selectively survey this large and growing literature next.

\subsection*{2.5.1 Overtrading}

Using data from a discount brokerage, Odean (1999), and Barber and Odean (2000) document that US retail investors’ portfolios exhibit very high turnover. Such excessive trading is a departure from standard models of portfolio choice, where households in efficient markets are expected to invest passively, or in more complicated models, use fairly simple rebalancing rules for funds spanning the space of risky asset returns. The possibility that households have significant investment acumen that justifies their observed levels of underdiversification and turnover has been considered carefully in both disaggregated and aggregate analyses. For example, Odean (1999), Barber and Odean (2000), and Grinblatt and Keloharju (2000) find that high turnover is negatively correlated with expected returns. Using more aggregate data, Campbell, Schwartz, and Ramadorai (2009) find that institutional trades forecast the direction of stock returns over longer horizons, though individual traders likely benefit from providing liquidity to institutions over shorter horizons (e.g., Kaniel, Saar, and Titman, 2008). Overall, the consensus in this work suggests that households incur unnecessarily high transactions costs, trading more frequently than is good for their financial well-being.

Behavioral mechanisms are commonly invoked to resolve the puzzle of excessive trading by households, with one leading explanation being overconfidence. This term has been modelled in the literature as traders overestimating the precision of their private signals—
which then leads to the over-responsiveness of their trading patterns to information arrival. We discuss these ideas in greater detail in the next subsection. While overconfidence has been explored in much detail, it is also worth noting that similar effects on trading can arise if traders have a tendency to downplay the precision of others’ signals, as in Banerjee, Kaniel, and Kremer (2009), and Banerjee (2011). Eyster, Rabin, and Vayanos (2019) term this tendency “dismissiveness,” and show that excessive trading volume can also arise if traders simply neglect (rather than dismiss) belief disagreements.

Static models with overconfident traders include Odean (1998) and De Long, Shleifer, Summers, and Waldmann (1991). In dynamic settings, such as in Daniel, Hirshleifer, and Subrahmanyam (1998) and Gervais and Odean (2001), traders are modelled as overconfident, and their behavior leads to overtrading that is particularly pronounced following unusual realizations of past returns. In Gervais and Odean (2001), traders extract biased signals from their past performance, and excessively weight their past successes when learning about their own trading skill. This results in overtrading, especially by younger traders, following significant (and often luck-induced) successes in the market. Similarly, Daniel, Hirshleifer, and Teoh (2002) propose a theory where investors’ confidence increases more in response to positive outcomes versus negative outcomes, and argue this theory can explain short-run momentum and earnings drift, among other market anomalies.

Barber and Odean (2001) show that the tendency to trade excessively is stronger for men than women, and also stronger for single than for married traders, and argue that this supports overconfidence-based explanations, as men are more prone to overconfidence than women. Grinblatt and Keloharju (2009) use data from mandatory psychological assessments of Finnish males and combine these with detailed trading data on these individuals. They show that those with levels of self-confidence in excess of performance in aptitude tests tend to have higher levels of trading volume. They make the distinction between overconfidence arising as a result of excessively tight variance around perceived private signal precision and that arising from an upwardly biased mean of perceived signal precision. Their results are more consistent with the latter mechanism—which they term the “better than average” effect. This literature is also related to an earlier literature on the psychology of gambling—for
example, Langer (1975) and Langer and Roth (1975) present lab experimental evidence that infusing a gambling context with some element of choice increases the subjects’ tendency to interpret random outcomes as reflective of their own choices (the “illusion of control”). Gambling behavior has also been linked to “sensation-seeking” behavior (see, for example, Biasczynski, Steel, and McCongaghy, 1997; Vitaro, Arnseneault, and Tremblay, 1999), and Grinblatt and Keloharju (2009) also find evidence consistent with this determinant of excessive trading in the Finnish data, correlating traffic violations with trading records.

### 2.5.2 Learning and Trading

A deeper investigation of household trading and portfolio choice also provides a window into how households learn in complex and dynamic decision environments like the stock market. Individuals’ excessive trading volume can be difficult to reconcile with Bayesian learning about the parameters of stocks’ return distributions, especially if investors are also rational about their expectations of transactions costs. Accumulating evidence in a wide variety of economic decision making contexts suggests that agents learn differently from the standard Bayesian characterization: in particular, agents seem to be influenced by both the signal and noise components of their past experiences (see, for example, Kuchler and Zafar, 2019; and Fuster, Laibson, and Mendel, 2010). The underlying causes of such non-Bayesian behavior, though, is still a subject of active investigation.

One important direction that the literature has explored is to better understand whether households exhibit a form of reinforcement learning, i.e., they pursue actions that have been rewarding in the past without distinguishing whether those experiences reflect signal or noise (Roth and Erev, 1995, 1998; Camerer and Ho, 1999; Crawford, 2013). For example, Charness and Levin (2005) find, using a carefully designed lab experiment, that subjects respond to a noise shock even when Bayesian updating would predict alternative actions.

In the household finance context, numerous papers have found that households appear to “learn” more than standard Bayesian updating would predict. This can happen because households simply extrapolate from recent data (e.g., Greenwood and Shleifer, 2014; Barberis, Greenwood, Jin, and Shleifer, 2018). This can also happen because households over-
 Yet another possibility is that agents draw more introspective inferences from random experiences, believing that chance outcomes are informative about their own ability. A large literature in the psychology of attribution (e.g., Haggag, Pope, Bryant-Lees, and Bos, 2019; Kelley, 1973; Bem, 1965; Fitch, 1970; Blaine and Crocker, 1993) and the psychology of gambling (e.g., Gilovich, 1983; Gilovich and Douglas, 1986) explores related ideas. Of course, agents could also learn more rationally about their own ability, a theme explored carefully using detailed trading data from Finland by both Seru, Shumway, and Stoffman (2010), and Linnainmaa (2011).

Anagol, Balasubramaniam, and Ramadorai (2019) explore these themes of learning using a unique dataset from India which combines random allocation of stocks in IPO lotteries with detailed trading records of retail investors both before and after the random allocation. By construction, differences in outcomes across treated investors—those randomly allocated the IPO stock—and the remaining control group solely arise from responses to noise rather than informative signals. They find that treated investors substantially increase their subsequent trading volume in the remainder of their portfolio if the IPO they are randomly allotted experiences gains, and they symmetrically reduce their subsequent trading volume in the face of losses on the IPO stock. They explore a number of theories and find that their result is most consistent with retail investors learning about their own ability from randomly experienced noise in a manner reminiscent of Gervais and Odean (2001); their evidence is also consistent with random noise attracting and focusing investor attention, thus creating a link between noise and subsequent outcomes (see, e.g., Gabaix, 2017; and Barber and Odean, 2008).

2.5.3 The Disposition Effect

An important finding in the literature on household trading is that retail investors exhibit a pronounced “disposition effect,” namely, a high propensity to sell winning stocks and
to hold losing stocks (Shefrin and Statman, 1985; Odean, 1998; Grinblatt and Keloharju, 2001b). The strategy is a losing one in markets with positive momentum, and runs contrary to predictions of tax optimization, which would recommend selling losing investments to benefit from the tax shield.\textsuperscript{32} Moreover, in markets in which the marginal investor exhibits the disposition effect, stock prices could substantially underreact to the arrival of negative news (e.g., Frazzini, 2006; Goetzmann and Massa, 2008; and Grinblatt and Han, 2005).

A leading class of explanations for this phenomenon is that households exhibit non-standard preferences. In early work, Kahneman and Tversky (1979) identify that in the lab, subjects demonstrate pronounced loss aversion relative to an initial reference point, seemingly weighing utility in terms of gains and losses (with losses “looming larger”) rather than in terms of final wealth levels. This then formed the basis for much work in prospect theory (Kahneman and Tversky 1979, 1992), which adds to gain-loss utility with loss aversion the additional assumptions of overweighting of low probability events, and convexity of preferences over the loss domain and concavity over the gain domain. This last assumption generates a tendency for agents to take risks in the face of losses, while behaving in a risk-averse fashion in the face of gains—thus justifying the empirically observed disposition effect (Shefrin and Statman, 1985; Gomes, 2005). Prospect theory and loss aversion have also formed the basis for more general behavioral models of portfolio choice and asset pricing (see, e.g., Benartzi and Thaler, 1995; Gomes, 2005; and Barberis and Huang, 2009).

Of course, to justify the range of empirical findings about the disposition effect, several auxiliary assumptions are necessary. For one, the disposition effect is often measured stock-by-stock, i.e., when measuring gains and losses, these are measured at the stock level rather than the portfolio level. To explain this, a common assumption is that households create separate mental accounts, evaluating (or narrowly framing) gains and losses at the security level rather than the portfolio level (see, for example, Shefrin and Statman, 1985; Kahneman and Tversky, 1979).\textsuperscript{33} However, accumulating evidence suggests that this assumption of

\textsuperscript{32}As many authors, including Odean (1998), and Campbell, Ramadorai, and Ranish (2014) document, the disposition effect dips towards the end of the tax year reflecting some propensity to realize tax losses. However, even accounting for this dip, there is a higher propensity to realize gains than losses.

\textsuperscript{33}Related to this, another possible explanation for overtrading is that households create a separate “mental
security-level mental accounts is too restrictive. For example, Hartzmark (2015) shows that individual investors are more likely to sell the extreme winning and losing positions in their portfolios, a phenomenon that he terms the “rank effect,” which suggests that investors conduct intra-portfolio comparisons. Moreover, a small but growing empirical literature finds that households’ experiences in individual securities have important spillover effects on the portfolio as a whole (An, Engelberg, Henriksson, Wang, and Williams, 2019; Frydman and Camerer, 2016).

A second issue is that the theory offers little guidance about the correct reference point from which households measure gains and losses. A standard assumption (Kahneman and Tversky, 1991; Weaver and Frederick, 2012) is that the appropriate reference price is the purchase price of the security, but Hirshleifer and Ben-David (2012) find no sharp discontinuity around the gain-loss threshold measured around this reference point (i.e., zero returns). Moreover, they find a higher propensity to sell big losers than small losers.

One approach to resolving this issue has been to consider increasingly complicated formulations of the reference point, culminating in the Koszegi and Rabin (2006, 2007) expectations-dependent reference point models. These models highlight that expected actions and reference points are endogenous to one another, and propose equilibrium concepts (personal equilibrium (PE), and preferred personal equilibrium (PPE)) that individuals mentally solve for their optimal reference point in the face of such endogeneity. Under this formulation it is not clear that we should expect a sharp discontinuity around the purchase price of the stock.  

Yet another potential solution is the “realization utility” theory of Barberis and Xiong (2009), which depends on investors experiencing a burst of utility at the time of stock sale, which is linearly related to the magnitude of the realized gain. Frydman, Barberis, Camerer, Bossaerts, and Rangel (2014) use fMRI imaging to test the underlying assumption account” for their overall stock market winnings (see, e.g., Thaler, 1985; Barberis and Huang, 2001), and increase their trading volume accordingly, in a manner reminiscent of the “house money” effect (Thaler and Johnson, 1990).

34We discuss this theory and its potential to rationalize the observed disposition effect in more detail in the next section.
of such utility jolts, and find evidence in support. While this theory does help to explain the absence of a discontinuity in selling propensity around the purchase price threshold, especially with auxiliary assumptions such as transactions costs and random liquidity shocks affecting investors, it does not explain the increasing tendency for households to sell in the face of large losses—the theory predicts a flat selling propensity in the domain of losses.

It is fair to say that the disposition effect remains a continuing puzzle. This also raises a broader issue in this literature, namely that the predictions of any theory set up to explain a given fact about household trading can often deliver contrasting predictions about other aspects of household trading behavior. This issue is exacerbated by the fact that the apparent hyperactivity of some households’ trading behavior in brokerage accounts is matched by the virtual inactivity of (these or other) households in other types of accounts, especially in retirement saving accounts, despite theoretical prescriptions on portfolio rebalancing. We discuss this issue next.

2.5.4 Inaction, Inertia, and Inattention

In contrast with the results on trading behavior described above, a growing body of literature suggests that households are less active than they should be when rebalancing their risky asset portfolios. An important example is that household behavior in 401k retirement accounts is best described as sluggish (Agnew, Balduzzi and Sundén 2003; Ameriks and Zeldes 2004; Choi, Laibson, Madrian, and Metrick, 2002, 2004; Madrian and Shea 2001). This finding of sluggish behavior is also echoed more broadly in household behavior in risky asset markets, and not just confined to behavior in retirement accounts. For example, Calvet, Campbell, and Sodini (2009a) show using comprehensive administrative data from Sweden that the aggregate household sector seems to demonstrate weak active portfolio rebalancing in risky asset markets, and perhaps more importantly, that this aggregate finding masks substantial heterogeneity in active rebalancing behavior and the dynamics of participation in the market. In particular, they find, consistent with much other literature in household finance on the incidence of household investment “mistakes” (see Campbell, 2006) that better educated and wealthier households are less prone to sluggish behavior than poorer and less
well-educated households. The finding of seeming inattention to rebalancing opportunities is echoed in Brunnermeier and Nagel (2008), Bonaparte and Cooper (2009), and Karlsson, Loewenstein, and Seppi (2009).

What explains this apparent sluggishness? Once again, strictly rational models have struggled to explain such apparent inactivity. One approach has been to apply canonical models of bounded rationality to explain household behavior. One approach is to assume that households have a limited capacity to process large volumes of information, and that such “capacity constraints” can affect their ability to instantaneously respond to incentives to take financial decisions (Sims, 2003). Another approach, which is similar in spirit, is to assume that households face “observations costs” of evaluating their portfolios when deciding on rebalancing strategies. Assuming that such observations costs are proportional to portfolio size, Abel, Eberly, and Panageas (2007, 2013) show that the optimal rebalancing rule is to observe and adjust the portfolio at periodic intervals, i.e., to follow a time-dependent rebalancing strategy. Alvarez, Guiso, and Lippi (2012) bring both theory and data to bear on this problem, and show that, in combined survey and administrative data on Italian households, households’ reported frequencies of observation and rebalancing of their portfolios do not strictly coincide. This runs counter to the predictions in Abel, Eberly, and Panageas (2007, 2013), as well as other models of portfolio rebalancing such as Duffie and Sun (1990) and Gabaix and Laibson (2000). To reconcile the theory with the data, Alvarez, Guiso, and Lippi (2012) consider a model in which physical adjustment costs for durable goods consumption co-exist with financial asset observation costs, and show that such a model delivers a combination of a time-dependent and state-dependent (i.e., depending on the agent’s level of financial wealth) portfolio rebalancing rules.

More recently, Pagel (2018) uses the Koszegi-Rabin (2007, 2009) concept to explain inattention, namely that investors could have “news utility” arising from changes in expectations about consumption, combined with a profound aversion to bad news. According to this

---

35Such a framework has been used by van Nieuwerburgh and Veldkamp (2009) to explain the “home bias” puzzle, i.e., the observations that investors fail to diversify internationally, concentrating their portfolios in their home country.
model, an investor’s observations of their portfolio depend on the tradeoff between observing the news (which causes a first-order decline in utility because of the kink around losses), and missing out on consumption-smoothing opportunities. When calibrated, the model delivers a low (roughly annual) frequency of portfolio rebalancing. Olafsson and Pagel (2018) use detailed data on personal finance management of a large set of Icelandic households to show that the frequency of account checking exhibits a profound asymmetry around gains and losses, supporting the news utility assumption in Pagel (2018).

Portfolio inertia is also related to the “endowment effect” documented in laboratory settings (Kahneman, Knetsch, and Thaler, 1991), in which random allocation of an object leads to a reluctance to trade it in exchange for another, i.e., having an object appears causally related to the agent’s willingness-to-pay for the object, and lowers willingness-to-accept for the same object. Loss aversion has also been proposed as a leading explanation for the endowment effect. Anagol, Balasubramaniam, and Ramadorai (2018) use the random allocation of IPO stocks to roughly 1.5 million Indian investors as a way to causally identify this effect in asset markets. They find that there is a pronounced endowment effect in these data, namely, that investors who applied for and were randomly allocated IPOs have a strong tendency to continue to hold these stocks while those who randomly didn’t receive the stock rarely purchase it in the future. Put differently, both treated and control groups exhibit inertial behavior with respect to the stock, with one group hardly ever selling it, and the other, hardly ever buying it, despite both putting money down for an allotment of the stock up front. They consider a range of explanations for their findings, and do not find evidence that their empirical results are well-explained by the Koszegi-Rabin (2007, 2009) models of reference-dependent preferences. They suggest that their results may be best explained by the possibility that receiving an endowment causes an agent’s attention to shift toward its more positive attributes (i.e., “warm glow” based explanations a la Morewedge, Shu, Gilbert, and Wilson, 2009). Bordalo, Gennaioli, and Shleifer (2012) present a formal theory of the endowment effect based on a similar mechanism, namely that ownership causes subjects to focus on the positive aspects of the owned good, and therefore makes them less likely to trade.
2.5.5 Overtrading or Inertia and Inaction?

Thus far, we have reported on two apparently conflicting literatures. One, based on discount brokerage data, demonstrates that households trade too often, thus incurring excessive transactions costs and damaging their overall realized returns. Another literature, based primarily on retirement accounts, finds that households hardly alter their portfolio allocations or indeed their instructions for the allocation of future payments into the account, over long periods of time. Is there time-series variation in the extent to which households are inertial or hyperactive? Is there substantial variation in the cross-section of households along these dimensions?

Some studies attempt to make progress on a more comprehensive characterization of trading behavior across all of the risky portfolios that households simultaneously manage, and on the intertemporal variation in household behavior across these portfolios. For example, Brunnermeier and Nagel (2008) show that households, by and large, appear to exhibit inertial behavior, barely rebalancing their risky asset share in response to changes in wealth. However they do provide some evidence that inertial behavior is weaker after households experience large capital gains, or after they experience large outflows from risky assets to cash. Calvet, Campbell, and Sodini (2009a) show that rebalancing behavior varies across households as well, and that wealthy, educated investors with better diversified portfolios tend to rebalance more actively than others. And Hoopes, Langetieg, Nagel, Reck, Slemrod, and Stuart (2016) study the period surrounding the Lehman bankruptcy using detailed U.S. tax records, and find that households in the very highest income groups responded with significant sales of stock that persisted throughout the financial crisis, and were more sensitive to changes in stock market volatility. Finally, Bilias, Georgarakos, and Haliassos (2010) examine two nationally representative datasets in the U.S. (PSID and SCF) in the periods running up to the burst of the internet bubble and shortly thereafter. They find that while inertia and inactivity are widespread, only between 10 and 18 percent of households own brokerage accounts and, of those owning such accounts, between two thirds and three quarters trade stocks within a year. They conclude that overtrading of stocks is not a general
tendency in the population, even though it is observed in brokerage accounts—which they
document are owned by a relatively small fraction of households and make up a fairly small
component of their overall portfolio of financial assets.

All in all, a unified characterization of investor trading behavior in risky asset markets
is an elusive but important goal. The picture that emerges from the mainly reduced-form
evidence produced thus far is one in which households are by and large hyperactive in their
directly managed portfolios, which are often significantly underdiversified accounts. More-
over, there is significant cross-sectional variation in the degree of such hyperactivity, which
is correlated with household demographic characteristics. To add even further complexity,
the overall averages hide the fact that there may be at least two distinct populations within
the broad group of households, with a smaller body of highly sophisticated traders making
high returns on their fast-moving portfolios mixed in with unsophisticated portfolio-churning
households who are clearly trading suboptimally. A further wrinkle in the body of evidence
thus far unearthed is that asset location matters: this over-activity in brokerage accounts
needs to be juxtaposed against significant inactivity in retirement saving accounts.36
Perhaps a different way to state the goal of a unified characterization of household behavior in
this literature would be to call for research leading to a carefully specified theoretical model
of a household operating across a range of risky asset markets with structurally estimated
preference and belief parameters that might depend on demographics. Such structural esti-
mation would ideally be conducted on a large and detailed dataset of a broad cross-section
of households making decisions over a significant time-span, and could help to reconcile and
synthesize the complexities unearthed in the evidence provided in the literature thus far.

Next, we conclude this section by discussing the large and important literature on house-
hold saving for retirement.

36 Whether the inactivity and overactivity exist simultaneously for the same household or capture two
distinct groups of households is an intriguing open question. This is because there is as yet no firmly
established empirical work that simultaneously tracks both the high-frequency and lower-frequency decisions
of households across all of their risky asset portfolios, combined with demographic information on households.
2.6 Saving for retirement

As populations around the world age, and countries progressively move from defined benefit to defined contribution (DC) pension systems, households come under increasing pressure to make sensible choices about saving for retirement. DC pension schemes have the advantage of being fully-funded and potentially allow individuals to choose their optimal saving rate as well as their optimal allocation of those saving. This flexibility should in theory be beneficial for households—the discussion in section 2.2.3. highlighted how both the optimal saving rate and the optimal asset allocation depend on multiple demographic characteristics, including age.\(^{37}\) That said, there is significant evidence that households have limited financial literacy and heterogeneous cognitive abilities, as described in sections 5.1 and 5.2 below. Lack of financial literacy or limited cognition can lead to private retirement accounts being associated with suboptimal outcomes for households, say because households fail to properly diversify (see, e.g., Ahmed, Barber, and Odean, 2018). Much of the literature, therefore, also focuses on the optimal design of DC schemes to facilitate better household choices, and has a strong normative element to it, as we describe below.\(^{38}\)

2.6.1 The importance of defaults

There is strong evidence that default options in retirement plans have a large impact on individual choices. Such default options include the decision to (automatically unless opted out) enrol individuals into the retirement plan at the point of commencing employment, as well as (automatically) setting the default contribution rate and the default asset allocation of those that choose to participate in the plan, unless they specifically choose to decide otherwise. Such auto-enrollment serves as a “nudge” for households (Thaler and Sunstein, \(^{37}\)Gomes, Michaelides, and Polkovnichenko (2009) and Dahlquist, Setty, and Vestman (2018) extend that framework by explicitly modelling both a liquid taxable account and a separate tax-deferred retirement account with separate portfolio decisions for both. In Gomes, Michaelides, and Polkovnichenko (2009) the contribution rate to the retirement account is also made endogenous.\(^{38}\) Another important fact is that pension designs are also often motivated by paternalistic and redistributive concerns. We do not discuss those here, since that would be beyond the scope of this article.

\(^{37}\)Gomes, Michaelides, and Polkovnichenko (2009) and Dahlquist, Setty, and Vestman (2018) extend that framework by explicitly modelling both a liquid taxable account and a separate tax-deferred retirement account with separate portfolio decisions for both. In Gomes, Michaelides, and Polkovnichenko (2009) the contribution rate to the retirement account is also made endogenous.

\(^{38}\)Another important fact is that pension designs are also often motivated by paternalistic and redistributive concerns. We do not discuss those here, since that would be beyond the scope of this article.
Madrian and Shea (2001) compare two situations. The first is a situation in which employees are automatically enrolled in a 401k pension plan when they are hired, and have to actively opt-out if they wish to do so. The second is one in which they have to make an active decision to enroll in the plan. They find that participation rates increase significantly under automatic enrollment. Furthermore, a substantial fraction of 401(k) participants hired under the automatic enrollment scheme choose both the default contribution rate and the default asset allocation, while very few of the others explicitly make those specific choices, often choosing to invest at lower rates. Choi, Laibson, Madrian, and Metrick (2003, 2004) find that automatic enrollment increases 401(k) participation rates in the firms that they study by about 50 percentage points at six months of employment tenure. Even at three years after employees join, the percentage of employees enrolled in the 401(k) plan is around 25 percentage points higher under automatic enrollment. In addition, under automatic enrollment, 65 to 87 percent of new plan participants choose the default contribution rate and invest fully in the default fund. As discussed in Choi, Laibson, Madrian, and Metrick (2003) defaults can influence behavior because opting out of a default is costly (and potentially time varying), and because individuals are prone to procrastinate, as in models with hyperbolic discounting (e.g. Angeletos, Laibson, Repetto, Tobacman, and Weinberg, 2001).\textsuperscript{39}

This behavior has important policy implications. Chetty, Friedman, Leth-Petersen, Nielsen, and Olsen (2014) show that the impact of alternative retirement saving policies depends on whether their implementation requires employees to take an active decision or not. Retirement saving subsidies only work if individuals actively respond to them, which often does not happen. In addition, most of those who do respond are largely just transferring other assets to their retirement accounts; these authors find that $1 of government subsidy only increases net saving by 1 cent. They also suggest, in contrast, that raising employer

\textsuperscript{39}Brown, Farrell, and Weisbenner (2016) document a positive correlation between survey-measured procrastination and the likelihood of choosing the default option in a retirement plan. Paradoxically, in a large-scale survey of plan participants, Choi, Laibson, Madrian, and Metrick (2001) find that 68% of them believe that their current saving rate is too low.
contributions increases retirement account balances without any action on the part of the employees, and has a significant impact on total wealth accumulation.

2.6.2 Other behavioral biases in retirement saving

Benartzi and Thaler (2007) document that contribution rates also appear to be influenced by simple heuristics—for example, individuals appear to contribute the minimum amount necessary to get the full employer match, or simply round to a multiple of 5 percent or some other salient percentage. Choi, Laibson, and Madrian (2009) present evidence for individuals exhibiting mental accounting behavior when choosing investment allocations with consequent impacts on investors’ total equity exposure. Benartzi and Thaler (2001, 2007) document “naive diversification”, in which individuals tend to divide their saving equally across a small number of funds, even when the asset composition of those funds differs—thus potentially leading to suboptimal asset allocations for those individuals. However, inconsistent with this view of menus influencing choices, Huberman and Jiang (2006) show that individuals do not divide their holdings (approximately) equally across all funds available in their retirement plans, rather choosing to divide their holdings across a small number of funds, regardless of the number of funds available in the plan. Finally, as discussed in section 2.4.7, Benartzi (2001), Mitchell and Utkus (2003), and Poterba (2003) show that individuals tend to invest a disproportionate fraction of their retirement saving in the stock of their own companies, which is at variance with standard financial theory about diversification, especially since individuals’ income risk is largely derived from the same source.40

2.6.3 Designing optimal defaults for contributions

If defaults are highly influential for behavior, then choosing the right default options becomes extremely important.41 This is a particularly difficult problem to solve. As discussed earlier,

40See Beshears, Choi, Laibson, and Madrian (2018) or Mitchell and Utkus (2004) for more detailed surveys of these and other biases.
41Some plans/countries have gone to the extreme of removing all elements of choice by imposing mandatory enrollment at a fixed contribution rate, making this discussion even more important.
the optimal saving rate and portfolio allocations depend on a large number of factors and vary significantly across individuals.\footnote{We postpone the discussion of the role of financial education in addressing these issues to section 5.1.2.}

Choi, Laibson, Madrian, and Metrick (2003) show that the theoretically optimal default contribution rate is highly sensitive to the distribution of optimal saving rates across employees. However, setting employee-specific defaults is often infeasible, since an employer cannot observe all employee characteristics that are relevant to these decisions. Furthermore, firms may also face legal ramifications by not treating all employees in the same way (Carroll, Choi, Laibson, Madrian, and Metrick, 2009).

In practice, firms with automatic enrollment schemes often have fairly low default contribution rates (2 to 3 percent), to avoid discouraging individuals from participating in the plan (Beshears, Choi, Laibson, and Madrian, 2009). However this concern might be overstated. Beshears, Choi, Laibson, and Madrian (2009) find that an increase in the default rate from 3 percent to 6 percent had no negative impact on the participation decision of new hires at the company they study.\footnote{However, they argue that the result should be interpreted with caution, since in this case 6 percent was the upper bound of a one-for-one employer matching contribution.} Carroll, Choi, Laibson, Madrian, and Metrick (2009) suggest an alternative approach, where new hires must make an explicit choice of whether or not to participate in the plan at the point of joining the company. By forcing them to make an active decision, employees are not allowed to procrastinate, but they are also not forced into a potentially inappropriate default. Carroll, Choi, Laibson, Madrian, and Metrick (2009) find that enrollment rates are still significantly higher under this scheme than when the default is an opt-out scheme. They point out that this “active decision” approach may be most appropriate when the underlying problem is procrastination, but if the important concern is employees’ lack of financial literacy, then a single default option might be preferred.

As shown in Figure 1, the typical household income profile rises over the life-cycle. Early in life, households are more liquidity constrained, and saving is primarily needed to insure against background risk. Therefore, it is not optimal to invest wealth at this stage of the lifecycle in a largely illiquid retirement account.\footnote{In some plans, certain “hardship events” allow individuals to withdraw without penalty, but this is not a} It is therefore optimal to build up retire-
ment saving more gradually, meaning that contribution rates should increase over time.\footnote{This point is made more explicitly in Gomes, Michaelides, and Polkovnichenko (2009), who consider a separate retirement account and solve for the optimal contribution rate.} Thaler and Benartzi (2004) propose a scheme along these lines, in which individuals commit in advance to a contribution rate plan where the rate increases as their salary rises—the ex-ante commitment element addresses the procrastination problem (as in Laibson, 1997).

2.6.4 Optimal asset allocation and default options for investment

While the contribution rate is tricky to determine theoretically, the optimal default investment in risky assets is easier to determine—a low cost broadly diversified index—and there has indeed been a significant trend towards offering highly diversified low-cost index funds as the default choice in most pension schemes. The difficult issue, however, is determining the fraction of retirement wealth that should be allocated to this index relative to the fraction to be invested in the riskless asset. Building on the intuition from Figure 3 that this ratio should change over the lifecycle, several DC pension plans now offer “target date funds” as their default option. These funds set their equity exposure in a manner that depends on the number of years to retirement for any given investor, and gradually decrease the risky asset allocation according to a pre-determined (linear) rule as time elapses.\footnote{Balduzzi and Reuter (2019) document significant heterogeneity in the average level of risk exposure across different target date funds.}

As equation (2) indicates, time-to-retirement only serves as a rough proxy for the ratio of present-value of financial wealth to labor income, suggesting alternative rules for target-date funds. For example, Dahlquist, Setty, and Vestman (2018) propose an improved default allocation that also takes financial wealth into account. To complicate the issue further, as discussed in sections 2.2.3 and 2.4.2, recent research suggests that the optimal equity share is a hump-shaped function of age—this in turn suggests a re-thinking of the optimal trajectory of target date funds’ risky asset allocation.\footnote{However these papers don’t consider that a separate retirement account is illiquid; arguably the more} Tax considerations are also highly universal rule, and those “hardship events” only cover a subset of the events that individuals want to insure against.
important—Black (1980) and Tepper (1981) point out that the preferred tax treatment of retirement accounts creates a tax arbitrage, so that investors should follow a pecking-order of asset allocation, by holding more heavily-taxed assets in tax-deferred accounts. They claim that investors should optimally invest liquid wealth taking into account tax-deferred account balances, which should be maxed-out with highly taxed assets to the extent possible.\footnote{48}

2.6.5 Adequacy of retirement savings

Several early studies find, using U.S. data, that even though some households are not well prepared for retirement, the majority of the population saves enough for retirement.\footnote{49} Engen, Gale, and Uccello (2005), and Scholz, Seshadri, and Khitatrakun (2006) estimate that households with adequate retirement wealth constitute 65\% and 84\% of the total eligible population, respectively, by solving life-cycle models and comparing the implied wealth accumulation with those of respondents in the HRS. Hurd and Rohwedder (2012) also use HRS data and conclude that 71\% of the individuals adequately prepare for retirement. Those papers assume that households make full use of their net housing wealth during retirement, which is a strong assumption given that in practice most households fail to do so, as we discuss later in this survey. Consistent with the importance of this assumption, the percentage in Engen, Gale, and Uccello (2005) falls to 56\% when only half of housing wealth is assumed available for use. More on this important topic can be found in Poterba, Venti, and Wise (2011), who provide a detailed analysis of the wealth composition of retirees.

In contrast with the picture found in earlier studies, Poterba, Venti, and Wise (2011b) study the wealth evolution of older retirees using the HRS, and report that almost half of conservative allocation early in life should primarily be to a more liquid account. Without extending the models to a setting with both accounts it is impossible to make specific quantitative predictions for optimal portfolios in DC accounts.\footnote{48}

\footnote{48}Dammon, Spatt, and Zhang (2004), Gomes, Michaelides, and Polkovnichenko (2009) and Fischer and Gallmeyer (2017) discuss how this result might change in models with additional features such as income risk, borrowing constraints and differential tax treatment of gains and losses.\footnote{49}

\footnote{49}See Poterba (2015) for a detailed discussion of the different approaches to evaluate the optimality of retirement saving.\footnote{47}
them died with virtually no financial assets, and that they lived exclusively on their Social Security income. This means that older retirees may therefore not be prepared to face any unusual or unexpected expenses had they lived longer. Munnell, Webb, and Delorme (2006) use data from the SCF and Purcell (2012) use data from the HRS to estimate income retirement replacement ratios, and both studies conclude that most households are not saving enough for retirement. Gomes, Hoyem, Hu, and Ravina (2020) use private data on more than 300 thousand U.S. workers, and conclude, looking at their current account balances as well as their saving and investment behavior, that the majority will reach retirement with insufficient wealth to maintain their current standard of living. In the same vein, Lusardi and Mitchell (2011b) find that fewer than one-third of those age 50 and above ever tried to devise a retirement plan, and fewer than one-fifth thought that they had been successful at doing so. They further document a strong association between financial knowledge and the likelihood of successfully completing a retirement plan.

Many of the studies that we survey in this area also underscore the substantial cross-sectional heterogeneity in wealth accumulation in the data (Castañeda, Díaz-Giménez, and Ríos-Rull, 2003). Angeletos, Laibson, Repetto, Tobacman, and Weinberg (2001) explain why a non-trivial fraction of the population does not save enough for retirement by modelling self-control problems using a hyperbolic discounting formulation (as in Laibson, 1997).

More generally, the more recent studies that we survey in this area tend to find less optimistic results about retirement saving adequacy and suggest that this will be an important area of work in the future. Naturally, the studies differ in (combinations of) methodology, assumptions, and data. However, the evidence suggests that the more pessimistic recent results might be down to negative changes in the economic environment more recently, as well as recent increases in cross-sectional income dispersion (Krueger and Perri, 2006). In support of such explanations involving changes in the economic environment or household decisions, Lusardi, Mitchell, and Oggero (2017) also document that recent cohorts have taken on more

50Similarly, Ahmed, Barber, and Odean (2018) show that, if we take into account the observed suboptimal investment behavior of households, then a large fraction of them are unlikely to meet their retirement income requirements if enrolled in a DC pension plan where they will make their own investment decisions.
debt early in life and as such, they are more vulnerable financially in retirement. Horneff, Maurer, and Mitchell (2019) also document a potentially large impact on retirement saving from a persistent low return environment. These results suggest that going forward, the financial situation of retirees might deteriorate even further, particularly in light of other emerging risks such as cuts to social security benefits and/or increases in medical costs.

2.6.6 Wealth decumulation

In section 2.6.5, we discussed that there is significant heterogeneity in wealth accumulation prior to retirement. If following retirement, financial wealth fell rapidly towards zero, as in Figure 2, such cross-sectional wealth dispersion should quickly diminish. However the evidence does not support this conjecture. Some households do indeed decumulate wealth rapidly, especially if they saved very little for retirement in the first place—Poterba, Venti, and Wise (2011a) document that nearly half of older retirees die with virtually no financial assets. Other households, however, save much more, and decumulate more slowly (Poterba and Samwick, 1997; Ameriks and Zeldes, 2004; De Nardi, French, and Jones, 2010).

In section 2.1 we also discussed several reasons why individuals might be reluctant to spend their saving quite rapidly at retirement. These include bequest motives (Kotlikoff and Summers, 1981; Dynan, Skinner, and Zeldes, 2002; Ameriks, Caplin, Laufer, and Van Nieuwerburgh, 2011), medical expenditure risk (Hubbard, Skinner, and Zeldes, 1995; De Nardi, French, and Jones, 2010; Yogo, 2016) and life-span uncertainty (e.g. Horneff, Maurer, Mitchell and Stamos, 2009; Cocco and Gomes, 2012; Yogo, 2016). As discussed in previous sections (2.4.1, 2.4.4, and 2.6.5), housing is by far the largest asset that individuals accumulate over their lives. In the absence of a bequest motive, it would be optimal for households to run down this source of wealth as well—this could be accomplished by downsizing to a smaller house, returning to renting, or by the use of reverse mortgages.\footnote{Reserve mortgages represent a particularly appealing alternative since they allow individuals to fully access their housing equity, while continuing to live in their homes until death and avoiding the necessity of moving late in life.}

However, very few households use reverse mortgages or decrease their home equity late in life.

2.6.7 Asset allocation and annuitization

As highlighted in section 2.2.3, optimal portfolio allocation during the retirement period varies substantially across individuals, depending on their levels of pension income, the strength of their bequest motives, other sources of non-financial wealth, and the different forms of background risk that they might face, such as potential health expenditures. As a result, the optimal risky share could be very high for some people, but very low for others, and it might increase or decrease with age depending on individual circumstances. Horneff, Maurer, and Stamos (2008), Horneff, Maurer, Mitchell, and Stamos (2010), Inkmann, Lopes, and Michaelides (2011), and Yogo (2016) do work on these important topics, carefully studying the asset allocation decisions of retirees.

Another important concern for retirees is longevity risk. Consider an individual with a life expectancy of 80. If she depletes her accumulated saving exactly at age 80, she has no financial wealth left in the case that she ends up living beyond 80. Conversely, if she dies before she reaches 80, she will have failed to use all of her saving. To deal with this risk, individuals can buy annuities, which are assets that provide a payoff while the policy holder is still alive. Annuities provide a perfect hedge against longevity risk, and in the absence of frictions, provide a higher return than comparable assets because of the implicit survival premium that is embedded in them. Yaari (1965) derives a set of conditions under which complete annuitization is optimal, a result generalized by Davidoff, Brown, and Diamond (2005) who show that, with complete markets and no bequest motive, if the return offered by the annuities to survivors is greater than the returns of otherwise matching conventional

---

52 Some annuities also provide income to a beneficiary after the policy holder has died.
53 In any given pool of annuity buyers, those who die do not receive a payment, so their benefits are distributed to the survivors. This is a form of cross-subsidization/insurance in the pool of annuity buyers.
assets, then full annuitization is optimal. However, despite these advantages, Friedman and Warshawsky (1990) and Brown (2007) document extremely low annuitization rates in the population, though Koijen and Yogo (2018) point out that the market for annuities has increased substantially in the last 15 years, and suggest that the low demand for annuities is less pronounced in more recent times.

Several explanations have been proposed for the low demand for annuities. One important concern is adverse selection, i.e., annuities are more likely to be bought by individuals who expect to live longer than average, and therefore, the premiums set by annuity providers must compensate for this risk (see Finkelstein and Poterba, 2004, 2014). Friedman and Warshawsky (1990) and Mitchell, Poterba, Warshawsky, and Brown (1999) confirm that annuities offer very low returns to investors, although the latter paper shows that these have improved over time. Other explanations include individuals’ bequest motives, as well as the fact that social security already provides individuals with a significant annuity value (Inkmann, Lopes, and Michaelides, 2011). In addition, full annuitization leaves individuals exposed to other risks, such as medical expenditure shocks (Ameriks, Caplin, Laufer, and Van Nieuwerburgh, 2011), and often does not offer inflation protection while also forcing individuals to lock in a riskless rate of return. This is a highly suboptimal portfolio allocation for the vast majority of households (Horneff, Maurer, Mitchell, and Stamos, 2010).

We now turn to the other side of the household balance sheet, and discuss household borrowing and liabilities in both secured and unsecured lending markets.

---

54 Horneff, Maurer, Mitchell, and Stamos (2009 and 2010) and Cocco and Gomes (2012) document the benefits of alternative annuity products in hedging longevity risk. For a discussion on the design and hedging of some of these products see Blake, Cairns, Dowd, and MacMinn (2006).

55 This discussion also means that (both life and health) insurance products become particularly relevant to the asset allocation decisions of retirees. We briefly discussed the demand insurance earlier in this survey.

56 Brown, Kling, Mullainathan, and Wrobel (2008) argue that the low demand for annuities also arises because they are often framed in a (risky) investment context, instead of as providers of a riskless consumption stream.
3 Borrowing

Households take on substantial secured and unsecured liabilities, and we discuss how households manage these liabilities in this section.

3.1 Mortgages

Mortgages are used to finance purchases of real estate, and they are typically the single largest household liability, because the house is the largest single asset on most households’ balance sheets (see section 2.4.4). We provide a few broad summary statistics which are useful to set the context for this section (see also section 2.3.5). First, as Badarinza, Campbell, and Ramadorai (2016) and Badarinza, Balasubramaniam, and Ramadorai (2019) show, mortgages are of first-order importance in both advanced and emerging economies, though mortgage penetration is substantially higher on average in advanced economies—this is likely, at least in part, as Badarinza, Balasubramaniam, and Ramadorai (2019) highlight, a result of the difficulties of collateralization of houses in emerging economies arising from the inadequacies of well developed systems of land title and registration. Second, internationally, mortgage takeup varies considerably over the lifecycle (Ramadorai and Household Finance Committee, 2017). In most advanced economies, mortgage market participation exhibits a characteristic hump-shape with maximum participation found at middle age (household head is between 35-55), and declining rapidly towards zero for households above the age of 65, consistent with (housing) consumption-smoothing intuition. Third, both participation in mortgage markets (as well as the fraction of total liabilities conditional on participation) is virtually non-existent for households below the median wealth level in most countries around the world (Ramadorai and Household Finance Committee, 2017). Fourth, mortgage arrangements vary substantially across countries (Campbell, 2013), with the US and Germany relying primarily on fixed-rate mortgages (FRMs); Australia, the UK, and many southern European countries relying almost exclusively on adjustable-rate mortgages (ARMs); and many other countries, such as Denmark, Sweden, the Netherlands, and India showing considerable variation in the relative prevalence of ARMs versus FRMs over time. Fifth, as Campbell (2013) and
Badarinza, Campbell, and Ramadorai (2016) highlight, this variation in mortgage contractual form is also accompanied by substantial variation in other contractual features, such as the prevalence of refinancing penalties, recourse provisions, and the funding structure—for example, the US has primarily government-supported securitization, while many European countries have adopted covered bond structures.

Research on mortgages has followed household decisions in all phases of a mortgage’s lifecycle, beginning with mortgage choice, continuing through to mortgage refinancing, and finally to understanding mortgage default. These are all complex decisions, and as with other areas of household finance, the literature continues to refine its understanding of households’ observed choices, which often appear to deviate widely from the prescriptions of standard theory. In some cases, this is because there is a developing understanding of households’ true constraints, while in others, this is because of households non-standard preferences and beliefs. We begin our survey of the literature by discussing work on mortgage choice.

### 3.1.1 Mortgage Choice

Mortgage choice is rendered particularly complex by the fact that the form of the mortgage contract has a number of different possible variations. One feature is how the interest, amortization, and payment are specified. “Interest only” and negative amortization mortgages can minimize mortgage payments and alleviate liquidity constraints for borrowers (Piskorski and Tchistyi, 2010), though they can substantially increase the interest paid over the lifetime of the mortgage. In environments in which there is a mortgage interest tax deduction, they are potentially optimal. However, a major risk for the borrower with interest-only mortgages is that there is a substantial refinancing risk arising from the “balloon payment” that is due at maturity, and for the mortgage provider, that there is higher default risk arising from the lack of (or even negative) amortization over the term of the loan, as can be seen in Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010).

Yet another important set of features is the size of the initial fees charged by mortgage providers at origination, and how these fees are traded off against future interest payments on the mortgage (Dunn and Spatt, 1985; Chari and Jagannathan, 1989; Brueckner, 1994; LeRoy,
Stanton and Wallace (1998) suggest that “points,” i.e., up-front payments in exchange for a reduced future interest rate, can be used by mortgage providers to screen borrowers on the basis of their unobserved moving probabilities, which explains the proliferation of mortgage contracts on offer to potential borrowers. Agarwal, Ben-David, and Yao (2017) show, however, that mortgage borrowers who take points tend to lose out by doing so, and tend to be less educated and financially sophisticated. Along the same lines, Liu (2019) shows evidence from UK data that is consistent with mortgage providers shifting towards issuing a greater number of high-fee products when faced with funding cost shocks, and Gambacorta, Guiso, Mistrulli, Pozzi, and Tsoy (2019) show evidence from Italy that banks engage in “steering” unsophisticated customers towards mortgages that banks prefer they hold.

This evidence is consistent with fees and mortgage features other than rates being “shrouded” attributes that financially unsophisticated households might not pay attention to while making decisions (see, e.g., Gabaix and Laibson, 2006; Carlin, 2009; Carlin and Manso, 2011). If consumers were effective at weeding out bad deals by searching efficiently, perhaps this wouldn’t matter as much. However, Bucks and Pence (2008) provide direct survey evidence that ARM borrowers are unaware of the exact terms of their mortgages, and Woodward and Hall (2010, 2012) study the fees that borrowers pay at mortgage origination, arguing that insufficient shopping effort leads to excessive fees. More recently, Bhutta, Fuster, and Hizmo (2019) use comprehensive new data from a mortgage industry pricing platform, and find a large gap between transacted interest rates and the same individual lender-market level “best deal” rate. The gap is larger for lower credit quality borrowers, but falls when rates rise, consistent with greater incentives for search in the presence of higher costs. The structure of intermediation is also important in this context—for example, Robles-Garcia (2020) studies the UK mortgage market, in which mortgage brokers advise households and receive commission payments from lenders. She finds that brokers increase competition by facilitating the entry of lower-cost lenders but that commissions distort brokers’ advice to households.

An important contractual feature that has been studied in detail in the literature has
been the choice of ARMs versus FRMs (Alm and Follain, 1984; Stanton and Wallace, 1999). Campbell and Cocco (2003) solve a dynamic model of the optimal mortgage choice of an investor facing a risky stream of labor income, and highlight an important set of tradeoffs between costs and risks that determines the choice. ARMs generate a constant total real payment over the life of the mortgage, but are subject to interim payment risk arising from inflation volatility. This risk can be offset by further borrowing, but can be painful in the presence of borrowing constraints. FRMs in contrast are subject to risk of a different kind—if inflation rises or falls, this affects the real value of the total repayment, with declines in inflation the principal concern. This can be offset by refinancing, but this imposes transactions costs on borrowers, and can be limited by restrictions on the borrower’s home equity and credit score.

Empirically, the evidence on mortgage choice has been focused on the models that homeowners adopt to assess the likely costs of ARMs vs FRMs over the lifetime of the mortgage. Koijen, van Hemert, and van Nieuwerburgh (2009) argue that homeowners estimate the average ARM rate over the likely tenure of a mortgage in an extrapolative fashion using recent data, to compare these estimated costs with the prevailing FRM rate. They are able to match time-variation in the US ARM share well using this approach. However, Badarinza, Campbell, and Ramadorai (2018) find evidence that homeowners appear to look one year forward when deciding upon an ARM versus a FRM using international panel data on mortgage rates and ARM shares, which supports the Campbell and Cocco (2003) argument that borrowing-constrained homeowners are more focused on the current ARM rate for current budget relief and the possibility of securing a larger mortgage in the presence of limits on current mortgage interest to income ratios.

Households’ mortgage leverage choices can have important consequences for financial stability more generally through their effects on overall household leverage (see, e.g., Mian and Sufi, 2016; DeFusco, Johnson, and Mondragon, 2020). The recent literature in this area has focused on the role of individuals’ beliefs about house prices and desires for housing consumption on their leverage choices. For example, Bailey, D’Avila, Kuchler, and Stroebel (2019) show that agents’ house price beliefs have an important effect on household leverage.
choices, with pessimism about house prices translating into higher leverage choices and lower downpayments, and vice versa. Benetton, Bracke, Cocco, and Garbarino (2019) investigate the extent to which equity financing schemes can replace mortgage financing, studying the UK government “Help to Buy” scheme for first-time homebuyers. They find that households used the funds from the policy to purchase more expensive properties, rather than to reduce leverage and house price exposure.

Household mortgage choices also have important implications for monetary policy transmission. Interest-rate pass through can be more effective if households take on ARMs, as reductions in interest rates affect household budgets directly, without the need for refinancing. Di Maggio, Kermani, Keys, Piskorski, Ramcharan, Seru, and Yao (2017) show that households with ARMs benefited substantially from (unanticipated at the time of mortgage choice) resets of their mortgage interest rates during the Great Recession, and that these resets have a causal impact on their spending on durable goods, an effect that is partially attenuated by household deleveraging. Moreover, they show that there are aggregate effects, in the sense that regions with high ARM shares recovered faster from the Great Recession. Cloyne, Ferreira, and Surico (2020) document, using survey data from the UK, that mortgage holders respond much more to interest rate reductions than renters or outright homeowners, and attribute this difference to the fact that mortgage holders are more likely to be liquidity constrained “wealthy hand-to-mouth” consumers a la Kaplan and Violante (2014). Interestingly, Cloyne and Surico (2017) also document that the consumption response of mortgage holders to tax changes is also more responsive than the response of households without mortgage debt, i.e., mortgage debt can also affect the transmission of fiscal policy.

3.1.2 Refinancing

Mortgage refinancing is an important issue in household finance. One theme that has been explored extensively in the literature is that households often fail to refinance when it is apparently in their interest to do so. This is perhaps not surprising given the computational complexity of arriving at the rational decision. There is generally a fixed refinancing cost to be incurred at the point of refinancing, and with stochastic interest rates, the optimal
refinancing threshold is the solution to an optimal stopping problem (see Chen and Ling,
1989; Agarwal, Driscoll, and Laibson, 2013). Moreover, in some countries, such as the US,
there are constraints on refinancing imposed on those with low creditworthiness, and those
in negative home equity.

The two possibilities impeding optimal refinancing (constraints and behavioral biases)
lead to very different policy recommendations, and there has been growing interest in at-
ttempting to uncover the relative importance of these factors. Johnson, Meier, and Toubia
(2019) and Keys, Pope, and Pope (2016) study samples of pre-approved refinancing offers
and find evidence to suggest that failures to refinance cannot entirely be attributed to con-
straints arising from declines in creditworthiness. Earlier attempts to control for the effects
of constraints include Archer, Ling, and McGill (1996), Campbell (2006), Caplin, Freeman,
and Tracy (1997), and Schwartz (2006).

Andersen, Campbell, Nielsen, and Ramadorai (2019) study data from Denmark, where
refinancing to lower the interest rate is virtually unconstrained. They find a high incidence
of failures to refinance (as do Agarwal, Rosen, and Yao, 2016, using US data), even in this
relatively frictionless setting. They contrast two potential sources of inaction in mortgage
refinancing and structurally identify them using a discrete choice model. The first is time-
dependent inaction, a la Abel, Eberly, and Panageas (2007, 2013), in which, despite it being
optimal to do so, households only refinance at periodic intervals, thus missing intermediate
periods of lower interest rates. The second is state-dependent inaction, in which households
do not refinance unless the incentive to do so is sufficiently high. They find evidence that
time-dependent inaction, i.e., seeming obliviousness to profitable refinancing opportunities is
mainly seen in older households with lower income, education and wealth. State-dependent
inaction, in contrast, is more often seen in middle-aged households with higher income and
wealth, who act as if their time is valuable, and only turn their attention to refinancing if
there is a sufficiently high incentive to do so.

Whether failures to refinance arise from the impacts of constraints or from more behav-
ioral factors, an accumulating body of evidence suggests that households’ failure to refinance
can significantly reduce the effectiveness of monetary policy in a recession (Auclert 2019;
Agarwal, Deng, Gu, He, Qian, and Ren, 2019; Beraja, Fuster, Hurst, and Vavra, 2019; Di Maggio, Kermani, and Palmer, 2020). Can governments help with policies to facilitate advantageous refinancing? Agarwal, Amromin, Ben-David, Chomsisengphet, Piskorski, and Seru (2017) study the US Home Affordable Refinancing Program (HARP), which relaxed housing equity constraints on refinancing by extending government guarantees on insufficiently collateralized mortgages refinanced by intermediaries. They find that the government was able to positively affect refinancing activity using HARP, but find evidence that competitive frictions in the refinancing market partly hampered the program’s impact. However, government policies can also make refinancing more difficult—DeFusco and Mondragon (2020) study an an FHA policy change that excluded unemployed borrowers from refinancing and increased others’ out-of-pocket costs substantially. They find that these changes dramatically reduced refinancing rates, suggesting that the imposition of constraints can significantly impede refinancing either directly, or through households’ increased psychological costs arising from bureaucratic impediments.

In a competitive mortgage market, sluggish refinancers essentially subsidize prompt refinancers through lender profits that are redistributed as lower interest rates, in a manner reminiscent of Gabaix and Laibson (2006). Increasing evidence suggests that banks manipulate such behavioral patterns. For example, Gurun, Matvos, and Seru (2016) show that subprime lenders sell more expensive mortgages in areas in which they advertise more, and that this effect is stronger when mortgages are sold to relatively unsophisticated customers.

3.1.3 Default

The literature has also explored the determinants of mortgage default rates. Of course, the dynamics of credit supply including screening and monitoring, are deeply important in this context, and this has been an important strand in the literature seeking to ascertain the causes of the mortgage crisis during the Great Recession (see Mian and Sufi, 2009; Keys, Mukherjee, Seru, and Vig, 2010; Purnanandam, 2011).

There has also been significant work on understanding how credit demand by households factored into the increases in delinquencies seen during the recent mortgage crisis. Mian
and Sufi (2011) identify a high elasticity of loan demand with respect to rises in home equity, and find that those with relatively low credit scores and high credit card utilization rates were primarily responsible for increased borrowing—and for the subsequent rise in defaults at the onset of the mortgage crisis. In contrast, Adelino, Severino, and Schoar (2016) highlight that mortgage originations in the pre-crisis period (2002-2006), in line with standard trends, were primarily dominated in dollar terms by middle- and high-income borrowers, and that mortgage delinquencies during the crisis were also primarily from such “prime” borrowers. They suggest that these findings are consistent with explanations based on the role of potentially misspecified house price expectations by households which resulted in defaults once prices dropped. Several authors explore the role of expectations about house prices, and provide evidence consistent with extrapolative behavior by households in this context, both from broad price indices as well as from personal experiences (Foote, Gerardi, and Willen, 2012; Cheng, Raina, and Xiong, 2014; Shiller, 2014; Glaeser and Nathanson, 2015; Kuchler and Zafar 2019; Armona, Fuster, and Zafar, 2019).

Another potential determinant of the rise in defaults arising from the household demand side is that households may not have completely understood the contractual features of their mortgages at the time of origination. While Mayer, Pence, and Sherlund (2009) suggest that such “exotic” features as negative amortization did not contribute much to the high rise in defaults observed during the recent mortgage crisis, Amromin, De Nardi, and Schulze (2018) do find evidence that default rates on “complex” mortgages spiked during the crisis despite such instruments primarily being taken up by wealthier, educated households.

What role do government regulations play in mortgage defaults? In the US, several authors have explored the role of subsidies to low income borrowers through the Community Reinvestment Act (see, for example, Dahl, Evanoff, and Spivey 2000; Kroszner, 2008; Agarwal, Benmelech, Bergman, and Seru, 2012). Using regulatory treatment discontinuities in India, Campbell, Ramadorai, and Ranish (2015) also find evidence that subsidies matter for mortgage default, and that when regulators force lenders to recognize delinquencies early in the life cycle of a loan, this can have significant impacts on longer-run default rates. Corbae and Quintin (2015) highlight the role of regulation in changing possible contract types, and
subsequently on default and foreclosure. Their model suggests that relaxations in payment to income requirements are an important determinant of defaults during the mortgage crisis.

Moving past the crisis, the literature has also recently begun to explore the impacts of new statistical screening technologies on the assessment of mortgage delinquencies, with attendant consequences for the distribution of mortgage credit to different groups of borrowers. Fuster, Plosser, Schnabl, and Vickery (2019) show that FinTech lenders have recently increased in importance as a source of loans in the US market, and that they process mortgage applications quicker than traditional lenders, without a large impact on defaults, suggesting that the new technology helps to improve mortgage lending efficiency, a finding echoed in Buchak, Matvos, Piskorski, and Seru (2018), who also find that such lenders seem to use different/non-standard information to set interest rates. Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2018) examine the performance of traditional and machine learning credit screening techniques estimated on large datasets of US mortgages, and find that while machine learning algorithms deliver greater predictive accuracy, predicted default rates for Black and Hispanic borrowers rise relatively more than those for White and Asian borrowers, and show using a structural model that this results in increases in rates for these groups in the counterfactual equilibrium in which all lenders adopt the new technology. Bartlett, Morse, Stanton, and Wallace (2019) show that controlling for observables, Black and Hispanic mortgage applicants face higher rejection rates than White applicants, but that decisions from FinTech lenders do not demonstrate such a pronounced gap.

Overall, the fast-growing literature on mortgages has continued to document the complexities of household decision-making in all phases of the mortgage lifecycle, from origination to refinancing to default. The substantial weight in both household assets and liabilities of housing and mortgages, respectively, makes household decisions in these areas fertile ground for extracting information about household beliefs, preferences, and constraints, and this is an active area of investigation, increasingly combining field experiments or insights from detailed administrative household-level data with “structural behavioral” models (see, e.g., DellaVigna, 2018). In the aftermath of the financial crisis of 2007-2008, research has also attempted to make progress on connecting household decisions in mortgage and housing
markets to broader macro outcomes such as financial stability and monetary transmission, and on exploring the potential that FinTech has to smooth (or exacerbate) frictions in mortgage markets. These questions deserve, and we believe will receive, deeper investigation going forward.

3.2 Credit cards and payday loans

Household credit-card borrowing has attracted considerable attention in the household finance literature, primarily because cards are an expensive source of credit, and because revolving card balances are often seen side-by-side with levels of liquid asset holdings. This phenomenon is difficult to reconcile with simple consumption-saving models.\footnote{Their role as payment instruments, in competition with debit cards and other forms of payment, is outside the scope of this survey.} A feature of credit cards that makes international comparisons very difficult is the extent to which they allow holders to revolve credit. In a number of countries, including the UK and the US, credit card holders are required to make minimum monthly payments, typically at 10% of the outstanding balance, but they can revolve the remaining outstanding balance at interest rates much higher than deposit rates. If credit-card debt is revolved in these countries, new charges incur interest from the time that they are made (i.e., there is no possibility of “floating”). In other countries, such as Germany, credit card issuers typically require automatic payment of the full outstanding balance at the end of the month. In the Netherlands, debit cards may be linked to a bank account that allows overdrafts up to a limit—this differs from a standard credit card, in that funds are withdrawn from the account at the time of purchase. This international institutional differentiation has pushed most of the existing household finance literature towards studying credit cards in the US and the UK, where a number of puzzling behaviors have been identified.
3.2.1 The co-holding puzzle

The “co-holding puzzle” (which is also known as the “puzzle of debt revolvers”) refers to the tendency of a significant share of credit card holders to revolve credit card debt balances, and at the same time have liquid assets that are more than sufficient to pay off this debt as well as to cover normal transactions needs. The puzzle was first documented by Gross and Souleles (2002), who used a large administrative set of US credit card accounts and found that this behavior is quite widespread among the middle class, and not connected to household financial distress. Such behavior has also been documented in the U.S. Survey of Consumer Finances, where around 30% of card holders that revolve debt were found to have liquid assets exceeding their outstanding balance. These liquid assets also exceed one-half the average monthly income of card holders; and more than two thirds of those card holders declared that they were usually not paying their balance in full, even though they typically wrote checks for the monthly payment (see Bertaut and Haliassos, 2006, who also review a number of proposed explanations). This behavior is costly, with median credit card annualized interest rates of around 15% over the period of study. Explanations for this co-holding puzzle have argued either that there are unobserved objectives (such as households seeking to maintain exempted assets prior to declaring bankruptcy) which would not be met by households moving funds from liquid accounts to the credit card account, or that the true service yield of outstanding debt or liquid assets to the holder is not fully reflected by the respective interest rates.

As an example of the former type of explanation, Lehnert and Maki (2002) argue that using existing assets to pay off debt may not be optimal for strategic debt defaulters, as (at least part of) these assets can be exempt if the household files under US Chapter 7 provisions. The authors utilize variation in the amount of bankruptcy exemptions across US states and find support for the hypothesis that co-holding is more pronounced in states with more generous exemptions. Gross and Souleles (2002), however, point out that co-holding entails interest costs and would be dominated by a strategy of borrowing on the credit card and purchasing exempt assets shortly before filing for bankruptcy.
An intriguing empirical finding of Gross and Souleles (2002) is that exogenous increases in the borrowing limit of credit card accounts tend to be met with additional spending by credit card holders, so that the prior utilization rate of the borrowing limit is restored within about five months. Bertaut, Haliassos, and Reiter (2009) develop an “accountant-shopper” model, in which the household member (or “self”) paying bills is matched with a shopper member (or “self”), that is also rational, but less patient than the “accountant” and therefore more prone to spending. The accountant realizes that any attempt to lower the outstanding balance by transferring funds from the liquid account will be met with new purchases by the impatient shopper. In this framework, co-holding preserves the liquid assets that would have been used up and is consistent with empirical evidence from the US. In a sample of UK data, Gathergood and Weber (2014) find that approximately 12% of households co-hold, on average, £3800 of revolving consumer credit, even though they could immediately pay it down. In support of the Bertaut, Haliassos, and Reiter (2009) model, co-holders are typically more financially literate, with above average income and education, and exhibit impulsive spending behavior.

Telyukova and Wright (2008) view the co-holding puzzle as an instance of the time-honored return dominance puzzle, dating back to Hicks (1935) who wrote: “So long as interest rates are positive, the decision to hold money rather than lend it, or use it to pay off old debts, is apparently an unprofitable one”. Telyukova and Wright (2008) argue, in the context of a monetary model with credit, that co-holding of return-dominated money and costly debt can be explained by the fact that cash is necessary to pay for some goods and services, for which credit cards are not accepted. The main issue is whether this explanation can fully account for the level of liquid assets observed under co-holding, given that careful self-control analyses also recognize the need for cash balances when specifying the threshold level of liquid asset holdings that is considered puzzling. Telyukova (2013) builds a dynamic stochastic heterogeneous agent model with a cash and a credit good to show that the model accounts for between 44% and 56% of the households in the data who hold consumer debt and liquidity simultaneously, and for 100% of the liquidity held by a median such household. She argues that, under reasonable calibration alternatives, the model can capture the entire
size of the “puzzling” group.

Individuals can gain or lose access to credit, and they can also experience unexpected fluctuations in their credit lines without losing access completely. A recent strand of literature invokes a risk that credit card debt revolvers face, namely that they repay (part of) their outstanding credit card debt and are then unable to secure access to an equivalent amount of credit in the future.\textsuperscript{58} In such an environment, the interest cost of outstanding debt balances should be set against the benefit of ensuring future access to a level of credit. Fulford (2015) documents changes in individual credit access at the extensive and intensive margins, and builds a model with fluctuating credit limits. A simulation of the model that incorporates observed credit limit variability generates a share of co-holding households at the level observed in US data (about one-third of the population). Druedahl and Jorgensen (2018) extend the buffer stock consumption model to incorporate long term revolving debt contracts, and find that co-holding is optimal for intermediate values of liquid net worth, even in the presence of a large wedge between debt and asset returns. When they introduce a risk of being excluded from new borrowing, correlated with the risk of unemployment, they can explain a large share of the “puzzling” group, and can also match a large portion of the distribution of credit card debt and liquid assets. Gorbachev and Luengo-Prado (2018) provide empirical evidence in favor of this explanation—their proxy for the perceived risk of future access to credit is based on whether the household has been denied access to credit, or was discouraged from applying for credit in the past, instrumented by the growth rate in the number of people that are served by a bank branch at the county level (this growth rate increases partly as a result of bank branch closures, which they argue affects access to credit). Although they provide evidence that a substantial fraction of the puzzling group acts rationally, and manage their future interest costs within manageable limits, they also find evidence that a smaller fraction of co-holders are significantly more likely to run into future financial distress and to declare bankruptcy.

\textsuperscript{58}US law gives financial institutions the right to refuse new credit, but not to demand immediate repayment of outstanding loans.
behavior is unlikely to be found, and that attention should probably be directed at assessing the relative contribution of various factors. Cross-country institutional variation is likely to be of great use in this context, both in assigning relative importance and in exploring the sources of the puzzling behavior by effectively “shutting off” certain credit card features and comparing outcomes.

3.2.2 The debt puzzle

Unlike the co-holding puzzle, which refers to credit card debt and liquid assets, the “debt puzzle” refers to holding substantial current credit card debt and at the same time accumulating considerable (illiquid) asset amounts for retirement. Laibson, Repetto, and Tobacman (1998, 2003) show that in a standard exponential discounting model augmented with credit cards, the degree of impatience needed to match observed credit card debt levels is not consistent with the observed high level of retirement wealth accumulation.

The authors instead propose an intertemporal model with hyperbolic discounting, where discount factors in consecutive periods, starting from the current period, take the form $[1, \beta \delta, \beta \delta^2, \ldots]$ . From the second period on, the household discounts each period relative to the immediately preceding one by a factor $\delta$. However, it acts more impatient with respect to the immediately subsequent period, since $\beta \delta < \delta$ . Taken together, in this model, the household acts more impatient with respect to the near future, and more patient with respect to longer-term objectives, both in absolute terms and relative to a standard exponential model. Importantly, behavior according to this model is dynamically inconsistent: a rational hyperbolic discounter knows that the “future self” will act more impatient as regards retirement accumulation, as retirement draws nearer, compared to what the “current self” would like. This creates an incentive for the current self to tie the hands of the future self, and it does so by buying illiquid assets (e.g., a house) that the future self cannot run down. Accumulation of illiquid assets can then co-exist with substantial credit card debt, because these assets serve the additional purpose of controlling the future self, while credit card debt is used for short-term consumption smoothing.

Meier and Sprenger (2010) elicit time preferences with incentivized experiments and
combine these elicited preferences with individual credit reports and tax returns. Subjects are asked to compare payments that are a given time apart (say, one month), but the distance from the present is varied. They find that present-biased (dynamically inconsistent) individuals are more likely to have credit card debt, and to have large outstanding balances, controlling for other characteristics.

Although the debt puzzle has been noted in the context of credit card borrowing, it involves a choice between current borrowing and asset availability over the longer term. As such, it is unlikely to be resolved by focusing further on special features of credit cards per se, as opposed to being cast in the broader context of the literature on behavioral biases.

3.2.3 Credit card debt literacy and differentiated contract terms

Credit card behavior of households faced with a variety of contracts and personalized contract terms has attracted the attention of researchers interested in assessing rationality of behavior, the tendency to make mistakes, firm behavior, and the effectiveness of regulatory responses. The multiplicity of contract terms, both within and across credit card contracts, provides fertile ground for testing a variety of hypotheses on all those dimensions, and the results of these studies have been quite remarkable.

Two key features of credit card contracts are the annual fee and the interest rate (APR). The tradeoff generated by these features has been exploited to test households’ ability to make sound financial choices. Agarwal, Chomsisengphet, Liu, and Souleles (2015) study an experiment involving the choice between two credit card contracts offered by a large bank: one has an annual fee but a lower interest rate than the other, while the second has no annual fee, but charges a higher interest rate on revolving debt. The approach that minimizes expected interest costs net of the fee is for households planning to be convenience users to opt for the contract with no annual fee, and for those planning to revolve debt to opt for the contract with the lower interest rate. The authors find that, on average, consumers choose the contract that ex post minimizes their net costs, but about 40 percent choose the ex post suboptimal contract, and some incur hundreds of dollars in avoidable costs. However, the probability of choosing the suboptimal contract declines with the dollar magnitude of
the potential error; and larger errors are more likely to be followed by a contract switch. Nevertheless, a small minority persists in holding substantially suboptimal contracts.

There is considerable evidence that households fail to respond optimally to differences in interest rates across multiple credit cards that they hold. Stango and Zinman (2009) find that, at the median of those in their sample who pay significant fees and credit card interest, almost half of credit card interest could be avoided by shifting balances from high to low-interest cards and repaying debt using available checking account balances. They use administrative data on every checking and credit card account transaction made by a sample of consumers over a two-year period, measure total explicit and implicit costs paid, compute avoidable costs, and examine how stable these two costs are over time. For most, the month-to-month credit card costs of both types are quite persistent. Most panelists could afford to consume more than $1,000 extra without any increase in the monthly interest payment.

Ponce, Seira, and Zamarripa (2017) in Mexico and Gathergood, Mahoney, Stewart, and Weber (2019) in the UK find substantial evidence for unresponsiveness of the relative repayments above the minimum required on all credit cards owned by the same individual, to the interest rate differentials between these credit cards. In the UK sample, the average difference in Annual Percentage Rate (APR) is 6.3 percentage points, approximately one-third of the sample average APR of 19.7 percent, but individuals allocate only 51.5 percent of their excess payments to the high APR card. Gathergood, Mahoney, Stewart, and Weber (2019) provide evidence that such behavior is most consistent with a suboptimal “balance-matching” heuristic.\(^\text{59}\)

The limited ability of households to optimize under differentiated contract terms leads to the question of how such terms are determined by the interplay of market forces and regulatory interventions. The main focus of this literature, on the interface between household finance and industrial organization, has been on credit card interest rates and annual fees, both in relation to the underlying cost of funds, and across customers with different credit scores. In a seminal paper, Ausubel (1991) pointed to the very limited reaction of credit card rates to changes in the cost of funds, spurring research on the relevance of search and

\(^{59}\)Some further examples can be found in Agarwal, Chomsisengphet, and Lim (2017).
switching costs (Brito and Hartley, 1995; Berlin and Mester, 2004; Galenianos and Gavazza, 2018). Analysis of variation in terms across customers has mainly emphasized asymmetric information and adverse selection (Stavins, 1996; Agarwal, Chomsisengphet, Mahoney, and Stroebel, 2015; Nelson, 2019).

A major regulatory reform, the 2009 CARD Act in the US, has been employed to assess effects on the distribution of fee-inclusive costs of credit card borrowing, the tendency of different customer segments to participate in the market, and (early) repayment patterns. The Card Act limited the ability of providers to adjust credit card interest rates or substitute other fees in response to unfolding information about customers, such as changes in credit scores or over-the-limit charges following origination. Agarwal, Chomsisengphet, Mahoney, and Stroebel (2015) employ a difference-in-differences strategy that compares consumer credit cards which were subject to the Act, to small-business cards which were not. They show that the Act reduced the average fee-inclusive cost of credit card borrowing, even though it raised the cost for higher-score customers that were pooled with others. Nelson (2019) builds, estimates, and uses a model of credit cards with switching costs across providers, adverse selection, and lender private information. Although he finds that average transacted prices do fall, there is exit from borrowing by customers who were previously facing the best terms within their credit score segment, reaching the 30% mark for some subprime categories. Nevertheless, he finds that the improvement in terms faced by consumers in greater need of borrowing results in an increase in market-wide consumer surplus; and in overall surplus (taking into account the fall in provider profits) among the prime borrower segment.

A promising direction for future research is to complete the circle: given that customers fail to optimize their credit card behavior to differences in credit card interest rates and other terms, do credit card providers take advantage of this failure? One possible avenue for providers is to make undesirable features of the credit card contract less obvious to customers, known as “shrouding”. A theoretical framework establishing the existence and persistence of shrouding equilibria in the presence of unsophisticated households was provided in an important paper by Gabaix and Laibson (2006). Ru and Schoar (2017) provide evidence that
credit card companies target less sophisticated customers with more back-loaded and hidden fees, opening the road to investigating relative takeup rates of such offers. Although hidden fees are one aspect of product complexity likely to disproportionately affect less sophisticated consumers, behavior that ignores interest rate differences, such as the “balance matching” heuristic, can also be very costly. Product differentiation provides an additional layer of “market-level complexity” on top of product complexity, making a regulatory emphasis on mere transparency a very inadequate response. Understanding the interplay of behavioral biases, product complexity, and product differentiation in different institutional (country) contexts will be important for shaping an adequate regulatory and policy response in the future.

3.2.4 Payday loans

Payday loans are a particularly expensive form of very short-term borrowing. Borrowers usually obtain a one-to two-week loan, and write a check post-dated for payment immediately after the next payday. Bhutta, Skiba, and Tobacman (2015) report that US payday lenders typically charge 10–20% interest, implying an annualized percentage rate (APR) between 260% and 1,040%. Payday loans have inspired a body of stimulating but inconclusive research on the motives of providers, the tendency of consumers to use them, their net effects, and the appropriate regulatory response.

Despite evidence that payday borrowers fail to exhaust available credit card lines before resorting to such a loan (Agarwal, Skiba, and Tobacman, 2009), there is considerable evidence that those in need have difficulty obtaining cheaper alternatives. Bhutta, Skiba and Tobacman (2015) report that nearly 80% of payday applicants have no unutilized credit card line, 90% have less than the typical payday loan amount ($300) available on credit cards just before applying, while 40% have no regular credit card. Unsuccessful shopping for cheaper mainstream credit surges around the first payday loan application when access to payday loans is restricted—Bhutta, Goldin, and Homonoff (2016) exploit variation in laws regarding payday loans in a difference-in-differences design and show that consumers, other than those with the lowest incomes, tend to shift to pawn shops rather than to credit cards
or consumer loans, when access to payday loans is restricted, leaving the overall number of	hose resorting to high-cost credit unaffected. From the supply-side perspective, payday loan
providers target households that are in need of short-term liquidity, however, neighborhood
racial composition does not have a significant influence on payday store location, controlling
for income, wealth, and demographics (Bhutta, 2014).

The study of payday loans is subject to data limitations. Although Gathergood, Guttman-
Keney, and Hunt (2019) are able to study 99% of approved payday loans in the UK and
their associated credit files over a two-year period, payday loans are not reported to US
credit bureaus. Researchers therefore tend to study data from a single payday lender, or to
focus on measures of access to such loans. Following Melzer (2011), some studies also ex-
plot within-state variation of geographical proximity to the border of states allowing payday
loans.

The effects of payday loans are highly controversial. Even if such loans cannot influence
credit scores directly, they can have indirect effects on financial well-being, households’ ability
to meet financial obligations, and ultimately credit records. On the negative side, Agarwal,
Skiba, and Tobacman (2009) find that, in the year following a successful application for
a payday loan, the probability of serious credit card delinquency doubles. Melzer (2011)
points to adverse effects on paying bills, Skiba and Tobacman (2019) to increases in personal
bankruptcy filing rates, Melzer (2018) to increased use of food assistance benefits, Carrell
and Zinman (2014) to declining job performance, Desai and Elliehausen (2017) to conflicting
effects on credit delinquencies across states and type of credit (revolving versus installment),
and Gathergood, Guttman-Keney, and Hunt (2019) to persistent increases in defaults and
violation of overdraft limits in bank accounts.

In contrast, a number of studies find positive effects of payday loans, as an otherwise
unavailable source of short term liquidity. Zinman (2010) finds that capping payday loan
rates leads to households reporting worse financial conditions. Morse (2011) finds that
payday loans mitigate about a quarter of the foreclosures and reduce larcenies in the year
after a natural disaster. Morgan, Strain, and Seblani (2012) show that banning payday
lending results in more bounced checks and complaints against debt collectors, while Zaki
(2016) reports improvements in expenditure smoothing between paychecks.

Yet others find no significant effect of payday loans. Bhutta (2014) finds little or no effect on credit scores, new delinquencies, or the likelihood of overdrawing credit lines. Bhutta, Skiba, and Tobacman (2015) reach similar conclusions with respect to household credit scores or delinquencies on average. These authors match administrative data from a payday lender to credit bureau data and base their regression discontinuity approach on people who were marginally rejected and those marginally approved for a payday loan, comparing the paths of traditional credit scores after the initial payday loan application. They find that the impacts on credit scores are close to zero using this approach.

A possible way out of this impasse is suggested by a recent paper Dobridge, (2018), who estimates, using data from the Consumer Expenditure Survey, a model of the form:

\[
Expenditure_{ict} = \beta_1 \text{WeatherEvent}_{ct} + \beta_2 \text{PaydayAccess} \times \text{WeatherEvent}_{ct} + \beta_3 \text{PaydayAccess}_{ct} + \beta_4 \text{Border}_c + \gamma W_{it} + \delta X_{at} + \theta Z_{ct} + \alpha_s + \alpha_t + \varepsilon_{ict},
\] (3)

regressing a consumption expenditure outcome on the occurrence of an exogenous shock (weather event), its interaction with a binary variable indicating whether the state of residence does not allow payday loans and the household lives in a county within 25 miles from the border of a state that allows such loans, the binary variable itself, a dummy variable for whether the household is within 25 miles of any state border, household, state and county controls, as well as state and time fixed effects. The coefficient of interest is estimated by Dobridge (2018) to be 0.2, indicating that access to payday loans mitigates the effects of exogenous shocks on consumption expenditure. These results point to “state-dependent” effects, i.e., following extreme negative weather shocks, access to payday loans moderates observed declines in overall spending, in nondurables spending, and in spending on food at home, mortgages, and home repairs. However, when extreme weather events are not present, access to payday loans causes households to report more difficulty paying rent, mortgage, and other bills.

On balance, the takeup of payday loans does not appear to be a sign of irrationality and behavioral biases, but rather a response to persistently restricted access to cheap sources of
very short-term liquidity. Especially at a time of rapid digitization, a natural way forward is to study ways in which such liquidity can be provided at much lower cost, and to understand the frictions impeding access to such funds. Understanding the business model of payday lenders will be relevant. Do the high costs being charged reflect true underlying risk and cost of funds, or instead, an attempt to exploit those in need or an inability for households to assess the APR implications of very short maturity loans? If conclusions point to exploitation, it is likely that provision of lower-cost alternatives should precede banning of payday lenders to avoid adverse consequences on households, and to facilitate competition in the market for short-term liquidity provision.

4 The Social Environment

4.1 Peer effects

Understanding how the behavior of peers influences household financial behavior is important for an analysis of how asset and debt participation spread through the economy, how good and bad practices become more widely adopted, and how financial crises can spread or begin through cross-household transmission. This is an active area of current research in household finance, and it is complementary to, but distinct from the literature on external habits and peer effects on consumption behavior that we do not review here.

The literature has so far focused on peer influences on asset and on debt behavior. Following the work of Manski (1993), peer effects are of three main types: endogenous effects, that arise from observing others use a particular product; exogenous or contextual effects, that are produced by characteristics of the peers, such as their education or financial literacy; and correlated effects, that arise from common characteristics of the individual and of the peers that induce similarity in behavior, or exposure of the individual and of the peers to common environmental factors, such as the number of banks or advisors in the area and the pattern of financial advertising. The household finance literature has taken an interest in the first two factors, while trying to control for influences of the third type.
4.1.1 Asset and debt behavior

On the asset side, seminal papers include Duflo and Saez (2002, 2003) and Hong, Kubik, and Stein (2004). Duflo and Saez (2002) study librarians allocated to different libraries in the same university and are exposed to different shares of colleagues participating in a new pension product. Their key finding is that participation in this pension product (and even the choice of mutual fund vendor) are positively related to the share of colleagues in the same department who use this product (or vendor), even for fine partitions of the possible peer groups, and using peer participation instrumented by the average wage or tenure in the department. Their analysis suggests the presence of endogenous peer effects in the use of these pension products, and in saving more generally. Duflo and Saez (2003) extends the analysis to an experiment of monetarily incentivized attendance at a seminar on tax-deferred annuities, where invitations and incentives were randomly allocated across departments and members.

Hong, Kubik, and Stein (2004) focus on the likelihood of social interaction. They find that households that report interacting with their neighbors, and attend church, are more likely to be stockholders, controlling for characteristics such as wealth, race, and education, as well as for personality traits such as risk tolerance and optimism. The authors do not observe characteristics of the peers, but they find effects to be stronger in US states exhibiting greater participation in stockholding. Brown, Ivković, Smith, and Weisbenner (2008) establish a positive relationship between an individual’s decision of whether to own stocks and average stock market participation in that individual’s community, instrumented with lagged average ownership of the states in which one’s nonnative neighbors were born. They are able to rule out correlated effects through the use of fixed effects and time-varying characteristics, and they find that the results are stronger in more sociable communities.

Kaustia and Knüpfert (2012) are able to match individuals to the peers located in the same zip code and find that positive recent stock returns on the aggregate stock portfolio of households in the zip code of an individual influence the individual’s decision to enter the stock market, especially in areas with better opportunities for social learning. Interestingly,
there is no symmetric discouragement effect for experiences of negative returns, suggesting that peers may only be sharing positive experiences.

Bailey, Cao, Kuchler, and Stroebel (2018) use data from Facebook to show that individuals, whose friends living in distant areas experienced house price increases, are themselves more likely to buy a house, and more willing to pay higher prices for the house they buy. They also provide evidence from a smaller survey suggesting that interaction with these peers raises house price expectations of individuals, even controlling for the correlation between house prices in the local area and those areas in which the peers are located.

On the debt side, Georgarakos, Haliassos, and Pasini (2014) exploit survey responses regarding peers to show that individuals who perceive themselves as earning less than the average of their peers have a higher probability of borrowing, larger outstanding debt amounts conditional on borrowing, and greater likelihood of running into financial distress. Bertrand and Morse (2016) show that non-rich households consume a larger share of their current income when exposed to higher income at the top of the local income distribution and provide indirect evidence that the non-rich may have relied on easier credit to finance this consumption increase. Agarwal, Mikhed, and Scholnik (2019) find that the size of a lottery win by one neighbor increases subsequent borrowing and bankruptcies among other neighbors. The neighbors become more likely to engage in visible consumption, and exhibit greater exposure to risky financial assets.

4.1.2 Inspecting the mechanism

Identification of the precise channel through which peers influence the financial behavior of households is a worthwhile challenge, especially with reference to issues of herding in financial markets and of the importance of relative standing. The modern literature on relative concerns and comparison effects dates back to Duesenberry (1949), but early ideas can be traced to Smith (1759). As discussed earlier, concerns about relative income or wealth

---

60 See Blume, Brock, Durlauf, and Ioannides (2011) for an extensive review of the issues related to identification of social interactions in different types of economic behavior. Models of herding incorporate learning from the behavior of others (Bikhchandani and Sharma, 2000; Chari and Kehoe, 2004).
status receive support in the literature on household debt. On the asset side, observing the positive returns of others may trigger the motive of “keeping up with the Joneses” (Abel, 1990; Gali, 1994; Campbell and Cochrane, 1999) encouraging groups of peers to adopt similar portfolios. However, relative wealth concerns may also encourage a tendency to differentiate one’s wealth prospects from those of others. Roussanov (2010) shows that when individuals strive to “beat the Joneses”, they have an incentive to opt for private risks, such as unlisted companies, where they alone will get richer.

When identifying peer effects, at least three important issues arise. First, the approach needs to pinpoint the source of contact that creates social interaction between peers. Second, there is a question of whether social interaction leads to imitation, information dissemination, or both of these channels simultaneously. Third, there is the question of the specific way in which observation, delivery, and processing of information occurs once peer contact has been established, and whether the peer effects are beneficial or harmful to households.

The literature has identified contact between peers in a variety of ways. Hong, Kubik, and Stein (2004) use self-reported sociability, measured through behaviors such as churchgoing. The literature on neighborhood effects is based on establishing geographical proximity between individuals, but this is often difficult to establish in household finance surveys or administrative data, as they are typically anonymized by removing locational information. Papers that employ locational information include Duflo and Saez (2002, 2003), discussed earlier; Kaustia and Knüpfert (2012), and Agarwal, Mikhed, and Scholnik (2019), who exploit zip code proximity; and Haliassos, Jansson, and Karabulut (2020), who observe the apartment(s) used by refugees over a twenty-year period after they are initially allocated.

One fruitful approach, to bypass the lack of locational information, is to include questions in household finance surveys regarding the perceptions that respondents have about people they consider to be their peers. While the identity and objective characteristics of peers

---

61 For a review of neighborhood effects on different types of economic behavior, see Durlauf (2004).
62 The Dutch National Bank survey contains a number of such questions (see, for example, Georgarakos, Haliassos, and Pasini (2014) on the link between debt and perceived relative income). Arrondel, Calvo-Pardo, Giannitsarou, and Haliassos (2019) include questions in the French TNS survey to elicit perceptions of respondents regarding their financial circle, their overall social circle, and the population.
continue to remain unknown, this approach has the advantage of using perceptions of the individuals regarding their peers, which also enter their decision making.

Another strand of literature focuses on using network information. Banerjee, Chandrasekhar, Duflo, and Jackson (2013) collect information from residents in 43 villages in rural India regarding their interactions right before a new microfinance product is introduced to them. They observe the precise individuals to which the microfinance company provides information, and are able to show that the “centrality” of individuals receiving the information (termed “diffusion centrality”) influences the size of the effect on villagers’ participation in microfinance schemes. Such tight control of information flows can be far more difficult in large asset or debt markets, such as the stock market, the housing market or the market for mortgages. One recent approach to surmounting this obstacle has been to use information from social networks such as Facebook to establish interactions among peers (see Bailey, Cao, Kuchler, and Stroebel, 2018; Bailey, D’Avila, Kuchler, and Stroebel, 2019). While both social interactions and housing outcomes are observed in these studies, the link between the two is not observed. The authors therefore supplement their analysis with a smaller survey dataset, to uncover a link between individuals’ social network interactions and subjective expectations.

The second issue, is whether social interaction leads to imitation, information dissemination, or both. More precisely, we would like to know the extent to which individuals learn from the choices of their peers (“social learning” or mindful imitation) rather than deriving utility from owning the same asset as their peers (“social utility”, “endorsement” or merely imitation). This issue often arises in the case of endogenous peer effects, when individuals and their peers adopt the same financial instrument simultaneously, possibly influenced by one another, but is also important in the presence of exogenous peer effects, or when trying to distinguish between the two. Consider exogenous effects such as those resulting from peers’ financial knowledge. If all of the effect runs through participating rather than non-participating knowledgeable peers, then we may well be mistaking an endogenous effect for an exogenous one.

Some evidence on this issue is provided by Banerjee, Chandrasekhar, Duflo, and Jackson
(2013) in their study of the social network in 43 Indian villages. They find that a microfinance participant is seven times as likely to inform another household in the network about the new microfinance product as a nonparticipant. However, non-participants are a majority in the data, and they account for about one-third of the total informedness and participation. Furthermore, given that a household is informed, its participation is not significantly dependent on the participation share of its peers, challenging the relevance of social utility. Further evidence is provided in Bursztyn, Ederer, Ferman, and Yuchtman (2014), who conduct a field experiment with a Brazilian financial brokerage and peer pairs. They ask one member of the pair to decide whether to place an order for a new asset, and they randomize whether that order will actually be filled. The authors also randomize whether the second peer learns anything about the first peer’s choice or outcome, only the choice, or both the choice and the final asset acquisition outcome. Comparison of the second peer’s choices in response to these three possible levels of information about the first peer allows estimation of the peer effect coming from knowledge that the other person wants to buy the asset (social learning) and any additional effect coming from knowledge that the other person actually holds the asset (social utility). The authors find that both effects are statistically and economically significant.

In order to circumvent the issue of endogenous selection of peers, Haliassos, Jansson, and Karabulut (2020) consider a refugee placement program in Sweden and study how the financial behavior of the refugees ten to twenty years after placement is influenced by the financial knowledge of their initial neighbors. They find significant effects of the share of initial neighbors who had business or economics education and some college enrollment on the medium-to longer-run participation of refugees in stocks and in private retirement accounts. This is not simply an instance of imitating participating peers: the estimated effect of the share of participants in the same asset is either insignificant or smaller than that of the share of financially knowledgeable neighbors, and the latter (mostly or entirely) persists even after participating neighbors are excluded. The presence and size of the effects depends on both the financial knowledge of neighbors and on the ability of refugees to process that information, as proxied by their education level. In another study, Arrondel, Calvo-
Pardo, Giannitsarou, and Haliassos (2019) exploit responses from a representative French sample regarding their perceptions of how informed or widely participating in stockholding progressively wider circles of peers are. They find that perceptions of greater information or participation among peers are related to more accurate stock market expectations of respondents, and that the effect runs solely through improved accuracy in perceptions of past stock market performance. They also find evidence for peer effects on stockholding behavior, in addition to those that run through expectations.

Girshina, Matthä, and Ziegelmeier (2017) reverse the question and study the influence of immigrant investment behavior on the stock market participation of natives, using data from Luxembourg, where about half of residents are foreign born. Immigrant stock ownership rates are instrumented with the corresponding participation rate in their country of origin. Contextual and correlated effects are separated from the endogenous peer effect by controlling for neighborhood-specific characteristics, individual risk preferences, and financial knowledge. The results show that peers’ investment attitudes have a sizeable effect on portfolios of the indigenous population.

Finally, the optimality of financial outcomes for those subjected to peer effects through information transfer, imitation, or both, remains a largely open issue. Shiller (1984, 1990) has pointed to naive extrapolation from the experiences of others, and this can be intensified further by selective reporting of good experiences of peers (Kaustia and Knüpfer, 2012). This issue links the literatures on peer effects with those on financial education and investment and borrowing mistakes—learning from peers can act as a substitute or a complement to learning from financial advisors, learning on one’s own, receiving financial education about complex financial products, or default options.

Overall, research on peer effects is challenging, given, among others, the issues of the endogenous selection of peers, and the well-known Manski “reflection problem.” particularly fruitful directions for further research include the study of alternative forms of information acquisition interact, either through complementarities or substitutability (e.g., peer information, that from financial advisors, and own research); the manipulability of peer input (e.g., through advertising, social media “influencers” or social networks); the possibility that peer
effects can grow explosively or “virally” (e.g., through recent advances in information and communications technology and the effect of social media); and finally, the issue of whether peer effects are more effective at surmounting inattentiveness and individuals’ cumulative cognitive burdens.

We now turn to documenting research on the interactions between culture and household financial decisions.

4.2 Cultural influences

There is voluminous literature on how culture influences the nature of institutions, and how institutions preserve culture (see Alesina and Giuliano, 2015). An important question is whether part of the differences in the economic environment facing households in their financial decision making can be attributed to differences in culture, over and above differences in policies and institutions.

North (1991) distinguishes between “informal” and “formal” constraints facing individuals: the former include “sanctions, taboos, customs, traditions, and codes of conduct”, while the latter include constitutions, laws, and property rights. Fernandez and Fogli (2006) describe culture as the “set of beliefs, perceptions, ideas, and collective experiences with institutions”. Guiso, Sapienza, and Zingales (2006) provide a framework for modeling the effect of culture on economic outcomes—while fast-moving aspects of culture are responsive to social interactions, they focus on slow-moving aspects. They envisage a channel running from culture, proxied by religion and/or ethnicity, through beliefs (priors) and preferences; or through political and institutional characteristics to economic outcomes. Beliefs and preferences include, for example, trust, risk aversion, and thriftiness.

Empirical work that links culture to household financial outcomes, as in other research areas, emphasizes the study of immigrants from different cultures faced with a common set of policies and institutions in the host country. The core econometric approach is to relate aspects of financial behavior, such as participation or conditional holdings, to a set of household characteristics and to an indicator of culture. This indicator is typically a country
dummy, or the average of the outcome studied that is observed in the home country, or some key institutional feature of the home country relevant for the outcome variable. These approaches share the assumption that household controls, including the proxy for culture, are linked to the financial outcome with the same, time-invariant coefficients, regardless of the immigrant’s origin. Decomposition methods have recently been used to allow for culture-and time-dependent coefficients to separately determine their effects on financial behavior.

Pioneering work on cultural influences in individual saving behavior by Carroll, Rhee, and Rhee (1994) does not find differences in saving patterns of immigrants to Canada by region of origin, but shows that recent immigrants, independent of origin, save less than indigenous Canadians and converge to the Canadian saving rate over time. Guiso, Sapienza, and Zingales (2006) look at national saving rates across countries and reach rather inconclusive results about whether they are influenced by the proportion of people in the country who think that thriftiness is important, instrumented by religious composition. Guin (2017) exploits the dramatic change in spoken language within short geographical distances in Switzerland to compare saving decisions of similar households on the German- and on the French-speaking sides of the language border, sharing a common institutional and economic environment. He finds considerably higher saving rates among German speakers, consistent with different distributions of rates of time preference. Osili and Paulson (2008) study immigrants’ stockholding behavior and find that it is influenced by the degree of protection of property rights in the home country, controlling for immigrant characteristics—this proxy is assumed to reflect the importance of home country culture.

Haliassos, Jansson, and Karabulut (2017) use two independent methods to group European countries into cultural groups: genetic distance as a proxy for past interaction across peoples (Cavalli-Sforza, Menozzi, and Piazza, 1994; Spolaore and Wacziarg, 2009); and the survey-based cultural dimensions of Hofstede (1984). They then assign the immigrant population into these cultural groups. The authors employ decomposition techniques that allow for time-and culture-dependent relationships between household characteristics and financial behavior. Controlling for differences in characteristics, they find significant differences in the propensity to own stocks, debt, and housing (i.e., of the primary residence) between
immigrant groups that do and do not share the same culture as the host country. They then compare the financial behavior of each immigrant group to that of the natives, controlling for differences in household characteristics, and find that the estimated differences decline gradually with exposure to the host country policies and institutions.

Exploring the channel through which culture influences household financial outcomes is usually done in two steps. The link between culture and preferences or beliefs has received considerable attention, but a proper review of this first step is beyond the scope of this survey.\textsuperscript{63} We focus instead on the work that links culture-induced beliefs or preferences to household financial outcomes. Because of its emphasis on the role of household beliefs or preferences as the main channel through which culture operates, this literature does not greatly rely on immigrant data, typically using microdata from a single country, and occasionally exploiting regional variation.

Guiso, Sapienza, and Zingales (2004) consider data from different provinces of Italy, and show that social capital, i.e., the advantages and opportunities accruing to individuals through membership in certain communities, encourages the use of basic financial instruments, such as writing a check, purchasing a share, and receiving formal credit, even after the individual has moved to another province. Social capital is thought of as a combination of social pressures and internal norms, and they proxy this using electoral participation and blood donation, two measures unlikely to be correlated with other features of the environment that are not part of social capital but could still influence the outcome. Alesina and Giuliano (2011a) find that reliance on family ties discourages civic engagement, political participation, and trust, all of which are relevant for building social capital. Guiso, Sapienza, and Zingales (2008) study the role of trust in stock market participation. They find considerable evidence that stockownership in a country is positively influenced by the extent of

\textsuperscript{63}Indicative examples include the role of religion in encouraging people to trust others more, trust the government and the legal system more, be less willing to break the law, and to be more likely to believe that markets’ outcomes are fair; the role of ethnic origin in influencing trust (Guiso, Sapienza, and Zingales, 2003 and 2006, respectively); and the role of culture in influencing preferences for redistribution (Alesina and Giuliano, 2011b; Luttmer and Singhal, 2011).
generalized trust exhibited by individuals in the country, which they model as the perceived probability of being cheated in the stock market. They also show that trust plays a role distinct from other concepts, such as aversion to risk, loss or ambiguity, as well as pessimism or low expected returns by utilizing detailed survey data.

The importance of research on the influence of culture in household financial behavior is likely to grow in the face of increasing household access to financial products across national borders and demographic groups, and as a result of large-scale migrant waves. In Europe, there are at least two important developments in this regard. The planned Capital Markets Union in the European Union, which aims at making location irrelevant for capital markets access, and the the recent refugee crisis, which implanted numerous people into institutions designed by a different culture. As societies become more multi-cultural, understanding the nature of cultural differences, and their persistence in the face of social interactions, exposure to common institutions, policies, opportunities, and information provided by social media is likely to be an important and fruitful area of further study.

We now move away from the broader environment to zoom in on households, next discussing the role of hereditary factors as well as intrahousehold analysis.

4.3 Hereditary factors

A strand of literature in household finance explores the “nature versus nurture” debate, using variance decomposition techniques to study the extent to which variation in financial behavior can be traced to genetic versus environmental factors. Cronqvist and Siegel (2015) find that about one third of the differences in saving propensities across individuals can be attributed to genetic factors. Although early conditions in the environment moderate genetic predispositions, there is evidence that the genetic component of saving is linked to genetic variation in time preferences, as well as the degree of self control. Barnea, Cronqvist, and Siegel (2010) use data on identical and fraternal twins and find that a genetic factor explains about one third of the variation in stock market participation and asset allocation, arguably linked to factors affecting effective participation costs, and risk preferences, re-
spectively. Cronqvist and Siegel (2014) attribute up to 45% of the variation in a number of investment biases to genetic differences, after controlling for observables. Such biases include under-diversification, overtrading, and the disposition effect, described earlier in this article. They find that personal experiences or events, as well as whether an individual has worked in the financial sector, additionally influence these biases. Cesarini, Dawes, Johannesson, Lichtenstein, and Wallace (2009) study experimentally elicited preferences for giving and risk taking and find that about 20% of the variation in these quantities can be attributed to genetic differences. Cesarini, Johannesson, Lichtenstein, Sandewall, and Wallace (2010) study fund choices made by Swedish households following a pension reform and attribute to genetic factors about 25% of variation in portfolio risk.

Data on twins has also been used to control for latent genetic factors and common family background, so as to more precisely estimate the link between factors of interest and financial behavior. For example, Grinblatt, Keloharju, and Linnainmaa (2011) use sibling data in their study of the relationship between IQ and stock market participation, and Calvet and Sodini (2014) study the wealth elasticity of risky portfolio shares, introducing yearly twin pair fixed effects in order to control for unobserved forms of heterogeneity in their data, such as attitudes towards risk, ability, genes, common family background, and expected inheritances. They find that the explanatory power of the twin pair year fixed effects, as well as the estimated contribution of genetic factors to variation in the risky portfolio share through variance decomposition, are quite responsive to the degree of interaction between siblings. These findings, together with the considerable contribution of observable characteristics in the cross-sectional variation of the risky portfolio share, support their argument that risky shares are not simply genetically determined, but rather, respond considerably to changes in financial circumstances and social interactions.

Overall, it is probably fair to say that existing research establishes the rather limited importance of hereditary factors for financial behavior. Barring major discoveries to the contrary, it seems fruitful to focus more on how the remaining portion of financial behavior is determined and can be improved, as we discuss in other parts of this survey.
4.4 Intrahousehold analysis

While virtually all existing household finance literature focuses on the household as the unit of analysis, there is a budding literature that seeks to understand the influence of factors affecting intrahousehold financial decision making. One approach taken by this literature is to incorporate in a life-cycle model of portfolio choice exogenous changes to family structure, arising from divorce, arrival of children, or the death of a spouse. Another approach seeks to understand the implications of intrahousehold bargaining for financial behavior, placing the focus on endogenous family transitions, possibly in response to exogenous shocks to the relative bargaining power of partners.

Taking the former approach, Love (2010) introduces exogenous shocks to both family status and the number of children, viewed as “background risks”, in a life-cycle model of consumption, asset allocation, and life insurance, subject to the additional background risks of labor income, out-of-pocket medical expenditures, and mortality. Younger households, which are mostly dependent on human wealth to finance consumption, tend to take higher risks with their portfolio, while the growing importance of financial relative to human wealth later in life tends to mitigate this risk exposure. Divorce affects the stock shares of men and women in opposite directions, widening the gender gap in exposure to stocks, but the presence of children moderates the predicted increase in stock exposure of working-age men. Marital status and children are predicted to have their largest influence at lower levels of wealth. Hubener, Maurer, and Mitchell (2016) expand this analysis by endogenizing labor supply and benefit claiming age, in a model with a detailed specification of US Social Security system rules. Married women are predicted to claim earlier than their single counterparts, whereas the opposite is true of married men, in the presence of a motive to hedge longevity risk. Because of non-separable preferences over consumption and leisure, the model generates a sharp drop in consumption after retirement, as well as claiming behavior that aligns well with observed data patterns.

The second approach views family transitions as the outcome of intrahousehold bargaining. Key to this approach is the modeling and measurement of the relative bargaining power
of the two partners. Relative income is an obvious candidate factor, even though others have also been considered. A number of paper focus on shocks to relative bargaining power, in response to which the partners choose whether to continue or break the partnership, and concomitant changes in portfolios occur. Addoum, Kung, and Morales (2016) introduce a limited commitment consideration in a life-cycle model of consumption and portfolio choice with cooperative decision-making by married spouses who have an option to divorce. Divorce takes place if both spouses obtain higher expected utility from exiting. Individual income shocks induce portfolio adjustments in the direction of preferences of the spouse who has gained relative power. Thus, overall household behavior exhibits time-varying risk aversion, conditional on whether the marriage survives. If the change in relative bargaining power results in divorce, then men adopt a riskier portfolio and women a less risky one than their (common) portfolio had been; and the opposite occurs for transition into marriage. These results are consistent, qualitatively, with predictions of models with exogenous family transitions, as well as with findings from PSID data: divorce leads to movements of the risky share in opposite directions, while increases in the wife’s relative income share tend to reduce household risk exposure. Also consistent with the implications of this model are the empirical findings of Addoum (2017), that transitions to retirement result in portfolio reallocations for married couples but not for singles: when men (women) retire, the household portfolio becomes less (more) risky, especially when differences in risk aversion between spouses are larger. Relatedly, Ashraf (2009) focuses on the process of bargaining and shows, in the context of a field experiment in the Philippines, that two factors that are relevant for outcomes are the observability of receipt of resources and the possibility for communication between the two partners prior to decision making. Olafsson and Thornquist (2018) take up a point made by Pollak (2005), namely that

---

64 These include relative non-labor income, relative age, entitlement to targeted transfers, legality of abortion, divorce laws, child benefits and alimony.

65 Regardless of gender, partners who are not the main financial decision makers in the household tend to deposit privately observed income to their own accounts; and to turn over jointly observed income to their partners when the allocation can be discussed with them. When discussion among partners is not allowed and the income is jointly observed, such partners tend to devote jointly observed income to consumption.
earning *potential* rather than actual income determines decision power. A zero or low current income of a partner who decided to take care of children may not be a good indicator, compared to a measure of access to resources that the partner would have if she or he decided to exit the cooperative marriage relationship and opted for maximum separation of finances or for divorce. The authors base their estimate of the gap in earnings potential between the two spouses on earnings of singles of the same age, gender, and education in the same region. To address the endogeneity of proposed measures of relative power, they instrument their gap measure by the part attributable to aggregate labor demand.\textsuperscript{66}

Consistent with the findings based on actual relative income reported above, the authors show that increased relative earnings potential of wives reduces the household propensity to participate in stocks and the conditional risky share, and it results in lower overall portfolio risk by reducing idiosyncratic risk.

Interestingly, there seems to be a lot in common in the predictions generated by the two approaches (i.e., the approach based on exogenous family transitions, and the one based on bargaining). The major challenge to the first approach is computational, once the unitary nature of the household is abandoned. For the second approach, it would be valuable to flesh out the nature of the bargaining process and tracing the effects of its key ingredients, but this will require data, observational or experimental, on the individual and joint financial behavior of partners. Increasing digitization of services is likely to increase observability of behavior, but privacy and confidentiality concerns need to be considered.

\textsuperscript{66}This is obtained by considering the industry composition of the county of residence of the couple, whether a given industry is male- or female-dominated, and national industry-wide changes in earnings (excluding the county where the couple resides. This approach is based on the assumption that the pattern of dominant genders in various industries is quite stable over time, and that differential earnings growth across counties is not immediately eliminated through labor flows.
5 Financial Literacy and Cognitive Abilities

5.1 Financial Literacy

Throughout this paper, we repeatedly refer to deviations of actual household behavior from the implications of standard rational optimization models. One possible source of such deviations, which has received considerable attention in the household finance literature since 2005, is lack of financial literacy.\footnote{For a more extensive survey of contributions in this field, see Lusardi and Mitchell (2014).} We first consider the relationship between measured financial literacy and financial outcomes, and then move to assessing the effectiveness of financial education interventions aimed at improving financial literacy.

5.1.1 Measured Financial Literacy

The seminal paper by Lusardi and Mitchell (2007) linked measured financial literacy to the tendency to save for retirement. Measurement was conducted using a set of special-purpose survey questions, whose answers were quite revealing and surprising. Among US households aged 51-56, 80 percent of the sample could interpret percentages, but only about half could allocate lottery winnings among five recipients, and 18 percent could correctly compute compound interest. In more recent studies, financial literacy is measured through three questions (known as the Big 3): one seeking to elicit understanding of compound interest, one to elicit whether households can perceive the difference between real and nominal interest rates, and finally, one on the benefits of diversification. Hastings, Madrian and Skimmhorn (2013) survey this literature, and show that a low fraction of households around the world exhibit facility with all three concepts simultaneously. An alternative approach to measuring financial sophistication is proposed by Calvet, Campbell and Sodini (2009b), who suggest a regression-based index that simultaneously utilizes information on three household “mistakes,” i.e., underdiversification, inertia in adjusting the risky share, and exhibiting the disposition effect.

Financial literacy has also been the subject of theoretical work. Delavande, Rohwed-
der, and Willis (2008) introduce costly acquisition of financial knowledge into a two-period model of portfolio choice between a riskless and a risky asset, and Jappelli and Padula (2013) present a two-period model and sketch a multi-period model of endogenous, costly acquisition of financial literacy, which implies endogenous co-movement of financial literacy and wealth over the lifecycle. They also point out that disincentives to accumulate wealth, in the form of generous social security systems, also discourage accumulation of financial literacy. Lusardi, Michaud, and Mitchell (2017) build a dynamic stochastic intertemporal model of consumption and portfolio choice with income and medical expenditure risk that explicitly introduces a process for accumulation of financial knowledge. Financial knowledge depreciates over time but can be augmented through costly acquisition of financial education. Financial knowledge leads to a higher expected return on the risky asset for an exogenously given amount of return risk. The model generates considerable wealth inequality, leading the authors to estimate that 30-40% of retirement wealth inequality can be attributed to endogenous financial knowledge. They conclude that imperfect knowledge creates substantial welfare losses for part of the population.

As mentioned earlier, the empirical literature has documented significant functional financial illiteracy around the world, as well as its greater incidence among women, minorities, and older people (Lusardi and Mitchell, 2008, 2011a). Other papers find a positive relationship between financial literacy and a wide range of “good” financial outcomes. These include, in addition to a greater tendency to save for retirement, a higher wealth level (Behrman, Mitchell, Soo, and Bravo, 2012; van Rooij, Lusardi, and Alessie, 2012; Jappelli and Padula, 2013), a greater tendency to participate in the stock market (van Rooij, Lusardi, and Alessie, 2011), a more limited tendency to use high-cost credit (Disney and Gathergood, 2013), a greater tendency to build diversified portfolios (von Gaudecker, 2015), a higher tendency to have a lower mortgage when young (Gathergood and Weber, 2017), and a lower tendency to report diminished spending capacity and limited available saving following a financial crisis (Klapper, Lusardi, and Panos, 2013). Instruments used to deal with the possibility of reverse causality and unobserved heterogeneity include self-reported mathematics scores at age 10, 68

---

68 This is later extended to allow for an additional effect of financial knowledge on risk diversification.
financial experiences of parents and siblings, opening of universities in the area, introduction of financial education requirements, and changes in educational policy during the subject’s formative years. Instrumental variables estimates have tended to be larger in magnitude than OLS estimates and to survive introduction of fixed effects in panel regressions, as well as of a variety of household controls. However the effect sizes have generally been estimated to be small, suggesting that more work is required to more clearly trace out these effects.

The abundant evidence on limited measured financial literacy internationally, and the observed relationship to financial outcomes naturally lead to the question of whether and how properly designed financial education interventions could positively influence household financial behavior. It is to this that we turn next.

5.1.2 Financial Education Interventions

There is a growing volume of literature on the effectiveness of financial education programs in various countries. Lusardi, and Mitchell (2014) review a variety of programs internationally and highlight their effectiveness in promoting financial literacy and positive outcomes. In recent research, Alan and Ertac (2018) conduct a randomized educational intervention among third and fourth graders in elementary schools, in which the students’ own teachers follow a structured curriculum in order to encourage patience in children. This is done by fostering, through case studies, stories, and in-class games, the ability to imagine future selves and the consequences of current actions, so as to promote self control and forward-looking behavior. Students exposed to this program exhibit increased willingness to defer consumption in incentivized experimental tasks: they require fewer gifts to defer consumption, and they allocate more consumption to the future when waiting is rewarded. Effects persist even after almost 3 years since the intervention, and they extend beyond financial outcomes, as treated students are less likely to receive a low behavior grade.

Bover, Hospido, and Villanueva (2018) study the effects of a 10-hour course delivered in compulsory secondary education, that covered budgeting, banking relationship and saving vehicles, as well as awareness about future financial outcomes. In a randomized field experiment among 9th graders, performance in standardized tests of financial knowledge increased
by 16% of one standard deviation, treated students were more likely to become involved in financial matters at home, and they showed a higher degree of patience in hypothetical saving choices. Treated students also exhibited greater patience than a control group, at various maturities and interest rates, in an incentivized saving task conducted three months after completion of the course.

A more skeptical approach to the effectiveness of financial education programs, at least as they are typically designed, has been taken by some authors. Willis (2011) has challenged the case for providing financial education on the basis of the costs a large-scale, mandatory program would entail, the loss in autonomy and privacy that individuals would suffer in a process of debiasing, and the relative cost effectiveness of alternative approaches, such as default options and regulation of financial advisor incentives. Fernandes, Lynch and Netemeyer (2014) present a meta-study, based on 168 papers covering 201 studies, of the relationship of both measured financial literacy and financial education interventions to financial outcomes such as saving, debt, investing, planning, and inertia. They find that financial education interventions explain only 0.1% of the variance of financial behaviors studied, and their role in low-income groups is even more limited. There are statistically significant but very small differences in effects across different types of interventions, such as counseling, financial education in high school, participation in seminars or workshops, and participation in a financial education program. Estimated effects tend to be smaller when a more demanding research design (randomized control trial) versus a quasi-experimental or pre-post comparison is employed. The meta-analysis finds bigger effects of measured financial literacy on outcomes, compared to those found for financial education interventions, raising the possibility that existing interventions have limited effects on subjects’ ability to identify a broad set of options and reach good financial decisions. The meta-analysis also finds that effects dissipate with time, and are negligible twenty months after the intervention. This leads the authors to argue in favor of “just in time” financial education tied to the specific behavior it tries to address and offered right before the financial choice is to be made.

A largely unexplored issue for the design and cost effectiveness of financial education interventions is how many people they ultimately reach and over what horizon. Building a
bridge between the literatures on financial literacy and on peer effects, Haliassos, Jansson, and Karabulut (2020) found that being placed next to neighbors with economics or business education, at least some of which was acquired in college, significantly increases the probability of participation in stocks and in individual retirement accounts ten to twenty years later. This social multiplier, however, is not operative when households on the receiving end do not have sufficient educational attainment to benefit from the information they receive, or sufficient exposure to such neighbors. In support of this evidence, Andersen, Campbell, Nielsen, and Ramadorai (2019) also find that having an economics or finance degree or a family member with an economics or finance degree leads to mortgage refinancing decisions that are closer to optimal.

The design and cost efficiency of financial education interventions, relative to alternatives such as product simplicity and default options, are likely to remain among the top policy issues facing governments in the years to come, as concern about wealth inequality grows and evidence that financial ability matters for returns accumulates. Hopefully, this literature will not get bogged down by the known difficulties in establishing the validity of instruments in IV estimation. It can, instead, leverage the growing potential for outreach provided by the internet, and employ treatments, such as information and nudging, to probe the factors that shape working financial knowledge of households and their ability to manage finances. If some categories, such as the young, can be reached more easily and effectively than others, it is also important to understand the workings of the social multiplier and to separately target those for whom the multiplier is inoperative.

5.2 Cognitive abilities

A potentially important factor, whose influence needs to be separated from financial literacy, is cognition. The relevant literature regresses asset or debt participation, or the tendency to make saving, borrowing or trading mistakes, on measures of cognitive skills obtained either from survey responses or from objective tests (typically from army enlistment procedures) and the usual set of household controls. This literature has been successful in establishing
statistically significant relationships between a wide range of financial outcomes and measures of cognitive skills. It has also found time variation in cognitive skills over the life cycle, as well as interactions between cognitive skills and other relevant factors, such as experiential knowledge or non-cognitive skills, that influence the ultimate quality of financial decisions.

A number of outcomes have been linked to cognitive abilities, conditional on other characteristics. Christelis, Jappelli, and Padula (2010) consider the tendency to invest in stocks directly or indirectly, through mutual funds and retirement accounts, and link it to measures of cognition, such as verbal, mathematical cognition and recall ability, among households aged 50+ in a positive direction. They interpret the weaker relationship for participation in bonds as consistent with cognition being related to information processing ability. Agarwal and Mazumder (2013) find that the probabilities of making mistakes in a particular use of credit cards or mortgage choice are closely related to the mathematics component of the Armed Forces Qualifying Test, but not to the verbal component. Gerardi, Goette, and Meier (2013) study subprime mortgage borrowers who took out loans in 2006 and 2007 and find that the probability of default on the mortgage is negatively related to the ability to perform basic mathematical calculations, controlling for a range of other characteristics, and that this does not arise from an improved choice of mortgage. Grinblatt, Keloharju, and Linnainmaa (2011) show that scores on IQ tests at mandatory army enlistment in Finland predict stock market participation, as well as trading behavior and performance.

Agarwal, Driscoll, Gabaix, and Laibson (2009) establish life-cycle variation in cognitive abilities and in the associated ability to make sound financial decisions. They study interest rates in six different markets, ranging from mortgages and home equity loans to credit cards, in order to understand the interaction between the decline of cognitive abilities with age and the simultaneous increase in experiential knowledge. The authors associate limited financial sophistication with a greater tendency to pay fees and higher interest rates and a more

69 They first consider the inappropriate use of credit cards for convenience transactions after a balance transfer, arising from the fact that, with an outstanding balance, finance charges start accruing from the time of purchase. Secondly, they study the tendency to accept a mortgage loan with worse conditions for failing to estimate the house value accurately, instead of reapplying elsewhere.
limited tendency to learn from one’s mistakes. They estimate that the resulting inverse-U-shaped financial sophistication over the life cycle peaks at age 53. Korniotis and Kumar (2011) find evidence supporting the hypothesis that older investors are more likely to use rules of thumb that reflect greater investment knowledge but exhibit less investment skill because they are less effective in applying their investment knowledge. This tendency is more pronounced among sensitive demographic groups, such as those with lower educational attainment, lower income, and minority status. While age effects provide a parsimonious way to interpret the findings, adverse cohort effects may also be at work among older age groups depending on the type of education they received and variations in this education over time.

Is there interaction between cognitive and non-cognitive skills? Lindqvist, Paues, and Vestman (2018) employ the introduction, in 2000, of a requirement that all Swedes born after 1937 choose how their individual defined-contribution pension account is managed. The relevant context is that this was combined with a campaign that encouraged Swedes to avoid taking the default option but did not provide any guidance as to optimal portfolio choices. The authors link high-quality, granular data on portfolio composition with cognitive and psychological test results from the military enlistment process. They find that those with higher non-cognitive abilities had lower returns, because they were more likely to opt out of the default option and to follow suboptimal portfolio strategies. By contrast, higher cognitive abilities were associated with higher returns, even conditional on opting out of the default option.

A possible avenue for further development of the line of enquiry into cognition would be to link up with recent research on the ability of the wealthy to generate, systematically and persistently, higher portfolio returns. The question is whether this should be traced primarily to the innate ability of the wealthy and, to some non-trivial extent, of their heirs (Fagereng, Guiso, Malacrino, and Pistaferri, 2020), to the willingness and ability of the wealthy to take more risk or to take it more efficiently (Bach, Calvet and Sodini, 2019; Campbell, Ramadorai, and Ranish, 2019) or to the investment of the wealthy in financial education (Lusardi, Michaud, and Mitchell, 2017). Shedding further light on how cognitive
abilities vary over the life cycle and interact with other factors, including financial education, could make an important contribution to the highly relevant debate regarding who is left behind, why, and what, if anything, can be done about it.

6 Financial Advice

Beyond financial education and default options, financial advice is another potential way to deal with differential financial literacy. It is also relevant even for financially literate people operating in a rapidly changing world of financial innovation and expanding availability of (cross-border) financial instruments.

The theoretical literature in this area focuses on defining and exploring the implications of conflict of interest between advisors and advisees. Inderst and Ottaviani (2009) build a theoretical model of financial advice in which two conflicts of interest and their interactions are studied. One is between the customer and the advisor, who both advises what is best for the customer and sells products. The other is between the advisor and the employing firm: the advisor faces incentives to sell, while the firm will eventually have to cover the cost of complaints and customer dissatisfaction. Inderst and Ottaviani (2012b) build a model of consumers, advisors, and product providers, with the latter offering hidden kickbacks or disclosed commissions, and argue that caps on commissions and mandatory disclosure can often have unintended welfare consequences. Gambacorta, Guiso, Mistrulli, Pozzi, and Tsoy (2019) build a model of banks selling mortgages, setting rates and providing financial advice to customers, some of who are sophisticated and others naive. Sophisticated customers know the mortgage suitable for them, while the others follow the bank's advice. They estimate a cost of distortion equivalent to an increase in the annual mortgage payment by 11%, but stress that banning advice is also costly, as even bad advice contains some information.

The empirical literature on the subject is also extensive. A first question, which dates back at least to Cowles (1933), is whether financial advisors know more than their advisees, or are at least less subject to the typical behavioral biases than those they advise.

\textsuperscript{70} For a review of the literature on financial advice, see also Inderst and Ottaviani (2012a).
Cowles warned against the inability of advisors to select superior stocks or to predict the future movements of the stock market. Womack (1996) finds that 14 major US brokerage firms were issuing “buy” or, especially, “sell” recommendations that were confirmed by the price drift after whatever initial movement might have been induced by the advice itself. However, new “buy” recommendations occurred 7 times more often than “sell” recommendations, leading the author to hypothesize ulterior motives, such as interest in future investment banking relationships, or the continuation of information flow from managers. Barber, Lehavy, McNichols, and Trueman (2001) find that timely adherence to consensus analyst recommendations requires high trading levels and corresponding transactions costs, leading to insignificant abnormal net returns. Bergstresser, Chalmers, and Tufano (2009) compare performance of mutual fund classes, defined according to whether they are sold to customers directly or through brokers. They find that funds sold through brokers offer inferior returns, even before the distribution fee, and that broker-sold funds exhibit the same return-chasing behavior as funds sold directly, with no superior aggregate market timing ability. Shapira and Venezia (2001) study brokerage clients of a bank and find that the disposition effect is present both for professional investors and for self-directed retail investors, but it is less pronounced among the former.

A second question that arises is the identity of those who are more likely to be matched with financial advisors. Hackethal, Haliassos, and Jappelli (2012) examine data from independent financial advisors and from a bank employing its own advisors. In both cases, advised customers tend to be wealthier, more educated, more experienced, and older. This is consistent with incentives linked to the volume of executed trades, but one should also keep in mind the issues of who wants to get advice and among those, who is likely to follow it. Collins (2011) and Finke (2013) argue that financial literacy and financial advice are complements rather than substitutes, partly because it takes financial knowledge to find good advisors. Bhattacharya, Hackethal, Kaesler, Loos, and Meyer (2012) study a random sample of active brokerage customers who were offered unbiased free advice. The investors who need the advice most are least likely to accept the offer. Even among those who accept advice, the majority do not follow it, although they estimate that following the advice would
improve performance, if the evidence of those who did is anything to go by. Mitchell and Smetters (2013) found that most households still rely on informal sources of advice.

A third question pertains to the type of advice given to households. Mullainathan, Noeth, and Schoar (2012) report results of a field study employing “mystery shoppers”, who made a first visit to a number of financial advisors pretending that they had money to invest and following a script regarding the amount and the current location of the funds. The advisors showed evidence of catering to the initial portfolio, as presented to them by the mystery shoppers, but also reversals during this first visit. They also exhibited a dramatic bias towards active management rather than index funds, as well as a tendency to differentiate advice depending on client characteristics, such as gender.

A fourth question is whether advice helps or hurts portfolio performance relative to what that performance would have been without advice. Hackethal, Haliassos, and Jappelli (2012) control for household characteristics and instrument use of an advisor with geographical controls. They find that advised accounts exhibit lower returns and Sharpe ratios, higher portfolio turnover, and greater diversification. Chalmers and Reuter (2018) use time-series variation in access to brokers to model demand for advice and infer counterfactual behavior. They show that, when brokers are unavailable, demand for target date funds increases differentially among new defined contribution plan participants with high predicted demand for advice. Clients of brokers earn significantly lower alphas and Sharpe ratios than matched target-date-fund (TDF) portfolios that provide similar risk and avoid broker fees. When brokers are available and TDFs are not, participants with high predicted demand for advice who invest through a broker are less likely to invest exclusively in a money market fund. On the debt side, existing research finds preliminary evidence to suggest that financial counseling can reduce indebtedness and delinquency rates.71

A fifth question is whether financial advisors engaged in financial misconduct are likely to disappear from the market. Egan, Matvos, and Seru (2019) track the universe of financial advisors in the United States from 2005 to 2015. Seven percent of advisors have misconduct records, with about one third of them being repeat offenders. Close to 50% of advisors lose

---

71See Lusardi and Mitchell (2014) for relevant literature.
their jobs after misconduct, but many get rehired by other, less desirable and less well-paying firms. Misconduct is more concentrated in counties with low education, older and wealthier customers, and this points to “specialization” in misconduct.

All in all, the existing literature on financial advice has raised serious concerns about the willingness of those most in need to obtain and to follow advice, the quality of the advice given and its dependence on incentive schemes, and the potential of market forces to mitigate financial misconduct. Yet, this is precisely the time when financial advisors can in principle be most useful in facilitating product access, even across borders, among those who are left behind or are subject to serious behavioral biases. The advent of robo-advice, with the lower costs it entails, could provide a key to democratization of finance and allow human advisors to spend more time talking to clients about the big questions of financial objectives, plans, and self-monitoring. A promising avenue for future research is to now move from diagnosing the problems and the perils of financial advice to designing the appropriate regulations and interventions, so that these can be overcome.

7 Conclusions

Household finance is an enormous and rapidly growing field. One testament to the richness of the field is the length of our (partial and focused) survey. The field studies the financial decisions of households, who operate in a complicated, difficult, and dynamic environment—an environment with a comparable or greater level of complexity to that faced by corporations. For example, households must initiate and refinance collateralized debt obligations such as mortgages and automobile loans, manage their credit quality and unsecured loan obligations over time, decide an optimal intertemporal consumption and savings plan, and manage assets to finance short-term requirements, as well as longer-run needs such as consumption in retirement.

Faced with these decisions, households often lack sophistication and training, and can face imperfectly competitive suppliers of financial products, who can exercise market power and exploit household errors in decision-making. The possibility of conflicts of interest,
inadequate governance, and imperfect regulation of financial intermediaries can exacerbate these issues. This complexity has given rise to many questions that need answering, leading to a rich literature. This is only the beginning—the field promises substantial growth and exciting work in the years to come.

That having been said, it is important for us to explicitly acknowledge areas that are extremely important to the field of Household Finance, but that we are unable to cover in our survey. First, while we discuss saving and borrowing decisions in detail, naturally, these decisions are also connected to households’ consumption decisions. Carefully documenting and exploring the determinants and implications of household consumption in the background of labor supply choices is a vast exercise in its own right, and fully deserves a separate survey. Second, when trying to understand household behavior it is crucial to realize that households are heterogeneous along many dimensions—not just in their circumstances and environments, but also in their preferences and beliefs. This is another area that we only indirectly reference in this survey, and we are unable to do justice to many aspects of this literature including the fast-growing literature seeking to understand the formation and persistence of household expectations about the future and perceptions of the past.

Third, some of the earliest research on household financial behavior, conducted long before the appellation “Household Finance” was introduced, emphasized (the lack of) portfolio diversification, even after accounting for more elaborate models of portfolio choice. Portfolio choice is another area that our survey touches only peripherally. Fourth, while we cover mortgages as a way to finance the purchase of a house, we reiterate here that we only superficially touch the immense and important field studying residential housing and renting decisions. Finally, but not least, we note that there is significant work on household finance conducted in policy settings and in finance practice, outside the coverage in purely academic journals. The field of household finance is both informed by and helpful for policy decisions and product innovations that seek to improve households financial lives and welfare, but we are unable to cover the excellent work and deep insights uncovered each day in policy, practice, and regulatory institutions. We have limited ourselves primarily to work published in academic outlets in this survey. We hope that all of these significant and important areas
and their relationship to household finance will be explored more fully in complementary future work.

As we conclude, we note four important and exciting developments emerging in the academic study of household finance. First, there is growing interest in the important question of how best to incorporate insights from household finance into asset pricing models, as well as both theoretical and empirical studies which seek to connect household micro-decisions to the likely effects of macroeconomic policies. This simultaneous development of links to Finance and Macroeconomics is not surprising, given that household finance traces its origins to researchers from both fields. We believe that the field can act as a catalyst in the process of mutual discovery that is currently ongoing in Macroeconomics and Finance.

Second, we note that a wide set of different tools is being deployed to tackle household finance questions, and that papers in this area frequently apply state-of-the-art approaches for empirical identification and inference and/or the solution and simulation of calibrated theoretical models. When these approaches are combined, they increasingly produce credible structural estimates of deep parameters that characterize household beliefs, constraints, and preferences, as well as frictions that households face—which in turn give us the ability to evaluate potential counterfactual scenarios.

Third, there is increasing interest in applying lessons learned from household finance to the design of regulation, both within and across international borders.

Last but not least, an important theme that is increasingly being explored is the promise and peril of technological solutions. Technology could hold the elusive key to many household finance problems, allowing for the possibility of “automatic” solutions. These might include the creation of platforms to easily allow households to switch financial products or providers thus reducing supplier market power, and the potential for remote help at low cost in the form of robo-advice or financial apps. However, technology may not immediately lead us to household finance utopia—there is much we do not understand about the potential dangers embedded in ease of access to customers and in biases associated with algorithmic decisions. There is also the possibility that ease of trading and switching results in damaging household hyperactivity, and finally, there may be issues of household trust in technological solutions.
that may impede their takeup.

Overall, we conclude that it is an exciting time to be working in household finance, and we eagerly look forward to further growth in the field.
References


Jackson & A. Bisin (Eds.), *Handbook of Social Economics* (pp. 93-131). San Diego, CA: North-Holland.


124(1), 86-112.


Bhutta, N., Skiba, P. M., & Tobacman, J. (2015). Payday Loan Choices and Conse-


111


113


87(1), 102-129.


Fagereng, A., Guiso, L., Malacrino, D., & Pistaferri, L. (2020). Heterogeneity and Per-


Grinblatt, M., & Han, B. (2005). Prospect Theory, Mental Accounting, and Momentum. 


Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The Ostrich Effect: Selective Atten-


Lusardi, A., & Mitchell, O.S. (2014). The Economic Importance of Financial Literacy:
Theory and Evidence. *Journal of Economic Literature*, 52(1), 5-44.


Peijnenburg, K. (2018). Life-Cycle Asset Allocation with Ambiguity Aversion and Learn-


Review, 80(2), 58-62.


Figures 1 plots an average life-cycle labor income profile estimated from the PSID. The particular profile is obtained using the estimation results from Cocco, Gomes and Maenhout (2005) with age-20 labor income scaled to 1.
Figures 2 and 3 plots the results from the baseline life-cycle portfolio choice described in section 2.2.4. Figure 2 plots average simulated consumption and wealth accumulation at each age, as well as the life-cycle income profile. As in figure 1, we have scaled age-20 labor income to 1. Figure 3 plots the average simulated share invested in stocks at each age until retirement.
<table>
<thead>
<tr>
<th>Issue</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>137 / 2019</td>
<td>Martin Kliem, Alexander Meyer-Gohde</td>
<td>(Un)expected Monetary Policy Shocks and Term Premia</td>
</tr>
<tr>
<td>136 / 2019</td>
<td>Luc Arrondel, Hector Calvo-Pardo, Chryssi Giannitsarou, Michael Haliassos</td>
<td>Informative Social Interactions</td>
</tr>
<tr>
<td>135 / 2019</td>
<td>Tiziana Assenza, Alberto Cardaci, Domenico Delli Gatti</td>
<td>Perceived wealth, cognitive sophistication and behavioral inattention</td>
</tr>
<tr>
<td>134 / 2019</td>
<td>Helmut Siekmann</td>
<td>The Asset Purchase Programmes of the ESCB – an interdisciplinary evaluation</td>
</tr>
<tr>
<td>133 / 2019</td>
<td>Josefine Quast, Maik H. Wolters</td>
<td>Reliable Real-time Output Gap Estimates Based on a Modified Hamilton Filter</td>
</tr>
<tr>
<td>132 / 2019</td>
<td>Galina Potjagailo, Maik Wolters</td>
<td>Global Financial Cycles since 1880</td>
</tr>
<tr>
<td>131 / 2019</td>
<td>Philipp Lieberknecht, Volker Wieland</td>
<td>On the Macroeconomic and Fiscal Effects of the Tax Cuts and Jobs Act</td>
</tr>
<tr>
<td>130 / 2019</td>
<td>Eduard Hofert</td>
<td>Regulating Virtual Currencies</td>
</tr>
<tr>
<td>129 / 2018</td>
<td>Olga Goldfayn-Frank, Johannes Wohlfart</td>
<td>How Do Consumers Adapt to a New Environment in their Economic Forecasting? Evidence from the German Reunification</td>
</tr>
<tr>
<td>128 / 2018</td>
<td>Christopher Roth, Johannes Wohlfart</td>
<td>How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior?</td>
</tr>
<tr>
<td>127 / 2018</td>
<td>Michael Haliassos, Thomas Jansson, Yigitcan Karabulut</td>
<td>Financial Literacy Externalities</td>
</tr>
<tr>
<td>124 / 2018</td>
<td>Athanasios Orphanides</td>
<td>The Boundaries of Central Bank Independence: Lessons from Unconventional Times</td>
</tr>
<tr>
<td>Jahr/Nummer</td>
<td>Autor/-innen</td>
<td>Titel</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>2018/122</td>
<td>Helmut Siekmann</td>
<td>Legal Tender in the Euro Area</td>
</tr>
<tr>
<td>2018/121</td>
<td>Maik H. Wolters</td>
<td>How the Baby Boomers’ Retirement Wave Distorts Model-Based Output Gap Estimates</td>
</tr>
<tr>
<td>2017/120</td>
<td>Helmut Siekmann</td>
<td>Die Einstandspflicht der Bundesrepublik Deutschland für die Deutsche Bundesbank und die Europäische Zentralbank</td>
</tr>
<tr>
<td>2017/119</td>
<td>Gregor Boehl</td>
<td>Monetary Policy and Speculative Stock Markets</td>
</tr>
<tr>
<td>2017/118</td>
<td>Gregor Boehl, Thomas Fischer</td>
<td>Can Taxation Predict US Top-Wealth Share Dynamics?</td>
</tr>
<tr>
<td>2017/117</td>
<td>Tobias H. Tröger</td>
<td>Why MREL Won’t Help Much</td>
</tr>
<tr>
<td>2017/115</td>
<td>Guenter W. Beck, Volker Wieland</td>
<td>How to Normalize Monetary Policy in the Euro Area</td>
</tr>
<tr>
<td>2017/114</td>
<td>Michael Binder, Jorge Quintana, Philipp Lieberknecht, Volker Wieland</td>
<td>Model Uncertainty in Macroeconomics: On the Implications of Financial Frictions</td>
</tr>
<tr>
<td>2017/113</td>
<td>Mewael F. Tesfaselassie, Maik Wolters</td>
<td>The Impact of Growth on Unemployment in a Low vs. a High Inflation Environment</td>
</tr>
<tr>
<td>2017/112</td>
<td>Gerhard Rösl, Franz Seitz, Karl-Heinz Tödter</td>
<td>Doing away with cash? The welfare costs of abolishing cash</td>
</tr>
<tr>
<td>2017/111</td>
<td>Jinhyuk Yoo</td>
<td>Capital Injection to Banks versus Debt Relief to Households</td>
</tr>
<tr>
<td>2017/110</td>
<td>Robert C. M. Beyer, Volker Wieland</td>
<td>Instability, imprecision and inconsistent use of equilibrium real interest rate estimates</td>
</tr>
<tr>
<td>2016/109</td>
<td>Helmut Siekmann</td>
<td>Replacing or Supplementing the Euro in Member States whose Currency is the Euro</td>
</tr>
<tr>
<td>2016/108</td>
<td>Helmut Siekmann</td>
<td>Restricting the Use of Cash in the European Monetary Union</td>
</tr>
</tbody>
</table>