The Gender Investment Gap over the Life-Cycle

Annika Bacher*

September 30, 2021

Click here for the most recent version.

Abstract

Single women are less likely to hold risky assets than single men and allocate a smaller share of their portfolio into stocks. This paper develops and estimates a portfolio choice model to quantify the determinants of the "gender investment gap" over the life-cycle. The framework allows for differences in household structure (single or couple), marital transitions as well as for rich gender heterogeneity along observable characteristics and stochastic processes. The model is able to rationalize the gender gap in equity shares and in asset holdings without introducing preference heterogeneity across men and women. Counterfactual simulations reveal that both current and expected lower income levels as well as larger household sizes of single women are the main determinants for explaining the investment gap. In particular, expectations about future income levels and household sizes drive most of the investment differences for young individuals whereas heterogeneity in current income levels (and household sizes) explains the gender investment gap later in life.

Keywords: Household Finance, Life-Cycle, Gender, Portfolio Choice

JEL: E21, G11, G50, J16

*European University Institute, Italy: annika.bacher@eui.eu. I am indebted to my advisors, Russell Cooper and Thomas Crossley, for their continuous guidance and support. Moreover, I thank my discussants Francisco Gomes and Oliwia Komada as well as Jesús Bueren, Dean Corbae, Mariacristina De Nardi, Eirik Eylands Brandsås, Philipp Grübener, Lukas Nord, Jirka Slačálek, and workshop participants at the ECB, the European University Institute, the University of Wisconsin-Madison, the University of Minnesota, the 2020 virtual ASSET conference, the 2020 European Winter Meeting of the Econometric Society, the 2021 RES Annual Conference, the 2021 GRAPE Gender Gaps Conference, the 2021 ESMES meeting and the 2021 EFA Doctoral Tutorial for comments and suggestions that helped to greatly improve this work. All remaining errors are mine. Part of this project was conducted while I was visiting the European Central Bank, whose hospitality is gratefully acknowledged.

1 Introduction

Single women are less likely to participate in the stock market than single men and if they do, they allocate a smaller share of their portfolio towards risky assets. However, in the presence of an equity premium and diversification gains, a more conservative portfolio translates (ceteris paribus) into lower wealth levels. This paper studies the sources of the so-called "gender investment gap" based on an estimated structural life-cycle framework. Generally, differences in investment behavior can arise due to differences in circumstances (such as income levels, income risk, number of household members etc.) or due to differences in unobservable characteristics such as preferences. In this paper, I ask how much of the gender investment gap can be explained by the former.

To that end, I first document life-cycle profiles of asset holdings, equity shares, stock market participation rates and equity shares conditional on participation ("conditional risky shares") for single men, single women and couples using survey data on US households who were born between 1945 and 1960. My empirical findings confirm the gender investment gap: Women are less likely to participate in the stock market and allocate – conditional on participating – a lower share of their portfolio towards risky assets. All differences are statistically different from zero, even after controlling for a wide range of observables such as age, education, labor income and the number of household members. In particular, the unexplained part of the gender investment gap decreases as households age.

To uncover which factors explain the remaining gap and to quantify the relative importance of each channel, I go on to develop a life-cycle model of portfolio choice that allows for exogenous marital transitions, for differences in household structure (single or couple) and in gender. In the model, single men and single women differ with regard to their income levels, their income risk, the number of individuals who live in their household, their marital transitions probabilities, the (expected) characteristics of their partner in the event of marriage, their survival probabilities as well as their out-of-pocket medical expenditures during retirement. I restrict preferences to be equal across single men and single women to study what fraction of the gender investment gap can be explained by differences in circumstances within the structural framework. However, I do allow for preference heterogeneity by marital status (i.e. between couples and singles) in order to better accommodate the data while at the same time keeping the model tractable.

In a next step, I estimate the model using data from the Survey of Consumer Finances (SCF) for financial choices and from the Panel Study of Income Dynamics (PSID) for labor income and demographic characteristics. To do so, I first estimate all parameters that can be cleanly identified outside of the model and afterwards estimate the remaining parameters using the Simulated Method of Moments (SMM), taking first-stage parameters as given. The model matches well the life-cycle profiles of wealth holdings and of equity shares for single men, single women and couples. Finally, I decompose the gender investment gap along the dimensions of gender heterogeneity within the model by replacing the female values with that of single men.

The main results are as follows. First, heterogeneity in labor income levels accounts for almost 40% of the gender gap in equity shares. Thus, the gender wage gap is amplified by translating into a less risky investment strategy, paying on average lower returns. Thereby, not only the current period income matters but also the fact that single men are endowed with more human capital (i.e. they expect higher income in future periods). Cocco, Gomes, and Maenhout (2005) show that labor income risk is uncorrelated to asset returns and therefore, it acts as a substitute for the safe asset. Consequently, a higher human capital endowment increases the willingness to take on financial risk for any given level of current labor income (and other state variables, such as wealth). Moreover, I find that differences in the number of household members are key in explaining the gender gap in equity shares. Over the course of their working life, larger female household sizes – which arises mainly through a higher likelihood of having children – can explain 43.16% of the observed gap. Again, not only the current household size affects savings and equity shares (through different consumption needs) but larger expected household sizes act as a future consumption commitment, making single women more vulnerable to financial shocks and hence, reduce financial risk-taking.

Lastly, I decompose the counterfactual scenario in which I assign single women the male income level into a *composition* and into a *policy* effect. The *composition* effect explains how much of the differences in female equity shares between the baseline and the counterfactual arises through changes in the distribution of individuals across the state space whereas the *policy* effect describes how much of these differences can be accounted for by changing policy functions for equity shares conditional on state variables. My findings suggest that in a world where single women had the same labor income level as single men, most of the increase in female equity shares early in life occurs because of the *policy effect*, that is conditional on states (through higher expected income in future periods). At around age 56, this relation flips and most of the increase in equity shares can be attributed to the distribution of individuals across the state space (as higher labor income in past years has translated into larger wealth levels). Hence, reduced form regressions that do not control for expectations about future income have less prediction power in explaining the gender investment gap early in life when these expectations are more important, which is in line with my empirical findings.

Related Literature. This paper contributes to several strands of the literature. First, it adds to a literature documenting gender differences in investment behavior and in financial choices. In general, there is large consensus that women invest less risky than men. For example, Sunden and Surette (1998) and Agnew, Balduzzi, and Sunden (2003) show that women in the US choose lower equity allocations in retirement saving plans than men. Similarly, Barber and Odean (2001) find that single men trade more often in risky assets than single women and attribute this result to male overconfidence. Säve-Söderbergh (2012) explores how men and women choose risk profiles in their pension contribution plans in Sweden. She documents that even though women do not less frequently include stocks in their portfolio, they do allocate a smaller share into risky assets. In more recent work, Almenberg and Dreber (2015) or Thörnqvist and Olafsson (2019) show that the gender investment gap in Sweden prevails until today. Ke (2018) attributes cross-country differences in stock market participation rates to gender norms, showing that countries with strong gender norms exhibit lower stock market participation rates of women. Moreover, Goldsmith-Pinkham and Shue (2020) provide evidence that women not only invest more conservatively in liquid financial assets but that they also earn lower returns on housing investments. Similarly, Andersen, Marx, Nielsen, and Vesterlund (2020) find gender differences in negotiation outcomes using transaction data of Danish residential real estate. My paper adds to this literature by being the first work – to be best of my knowledge – to analyze the gender investment gap in a structural framework. Most of the previous literature focuses on measuring the gender gap whereas my setting allows to model different channels and to quantify their relative importance. Relatedly, there exists an experimental literature that documents a higher risk aversion for women, also with regard to financial choices (see e.g. Croson and Gneezy (2009) for a review). Even though my model replicates differences in equity shares with homogeneous preferences across gender, my results are not in contrast to previous experiments: Single women behave *observationally* different to single men in the model, conditional on observable characteristics. However, my structural frameworks reveals that it is not underlying preference parameters that drive this result but rather expectations about lower income levels and larger household sizes (which act as future consumption commitments) in future periods.

Second, my paper relates to a literature that explores how family related shocks (such as marriage or divorce) affect portfolio allocation and savings. Cubeddu and Ríos-Rull (2003) study the role of marriage and divorce on wealth accumulation in a dynamic setting. Love (2010) was the first to present a joint life-cycle framework of marital status and portfolio choice. He finds that married investors hold more risky portfolios than singles. In the event of divorce, stock holdings increase for men whereas they decline for women. Hubener, Maurer, and Mitchell (2015) extend the analysis by incorporating endogenous labor supply and realistically calibrated social security benefit claiming. Again, they show that the equity share of couples is higher than for singles and that uncertain fertility can significantly reduce the amount of stock holdings. Christiansen, Joensen, and Rangvid (2015) empirically address the heterogeneous impact of family shocks on portfolio choices across gender with a difference-in-difference approach using an administrative panel dataset from Denmark. Similar to Love (2010) for the US, their findings suggest that the fraction of risky assets in women's portfolio increases after marriage whereas it declines after a divorce. For men, this relationship points in the opposite direction. Along the same lines, Bertocchi, Brunetti, and Torricelli (2011) find in an empirical framework that the marital gap of stock holdings (i.e. that stock holdings tend to be higher for married than for single individuals) in Italy is larger for women than it is for men. However, while all these papers show that family-related shocks affect portfolio choices heterogeneously across gender, neither of them quantifies the importance of such for the gender investment gap over the life-cycle. Finally, Bogan and Fernandez (2017) find that having a child with mental disabilities decreases stock market participation but increases the share of wealth allocated to risky assets, conditional on participating.

More broadly, my paper extends a literature that studies life-cycle pattern of household finances. Typically, life-cycle models of portfolio choice predict the optimal equity share to be decreasing in the ratio of current financial wealth over the present value of human capital (Merton (1971), Viceira (2001)).[1] Consequently, it should be optimal for young investors (who are endowed with relatively little financial wealth compared to their human capital stock) to allocate 100% of their financial wealth into stocks and to decrease the equity share as they age. In contrast, we observe only limited stock market participation and (conditional) equity shares, especially for young investors, in the data. The literature has proposed several mechanisms to explain this discrepancy. The most prominent explanation are costs associated with stock market investment (Vissing-Jorgensen (2002), Gomes and Michaelides (2005), Alan (2006)). Moreover, some papers have highlighted the importance of the illiquid nature of housing (Cocco, 2005), lack of financial literacy (Lusardi and Mitchell, 2014) and cyclicality of labor income (Catherine, 2019). However, so far, little focus has been on marital status as a source of financial risk that limits the propensity of investors to take on additional risk in the stock market.

Finally, my paper is methodologically related to Cooper and Zhu (2016) as they too estimate a life-cycle model of portfolio choice to study why certain subgroups of the population display different investment patterns. In contrast to the present study, the focus of their paper is on education and not on gender. Their findings suggest that income levels are the major deter-

[1] The intuition behind this finding is that the correlation between labor income (human capital) and asset returns is almost zero in the data and therefore, human capital acts as a substitute for safe assets (Cocco, 2005).

minant why more educated households invest more heavily in risky assets. Similarly, I find that income differences across single men and single women explain a significant part of the gender investment gap. With regard to my modeling approach, this paper is one of the first to introduce couples, single men and single women into a unified portfolio choice framework. I follow Fagereng, Gottlieb, and Guiso (2017) in their way to introduce a portfolio choice and build on Borella, De Nardi, and Yang (2019) to introduce exogenous marital transitions and the assortativeness of couple formation.

The remainder of the paper is structured as follows. Section 2 presents empirical observations on gender specific portfolio choices. Section 3 introduces the structural model. Section 4 introduces the estimation strategy and Section 5 presents the quantitative results. In Section 6, I decompose the gender investment along different sources of gender heterogeneity in the model. Finally, Section 7 concludes.

2 The Gender Investment Gap in the Data

The following section first explains the data and the sample selection criteria. In a next step, I provide empirical evidence on portfolio choices of single men, single women and couples over their life-cycle.

2.1 The Sample

Throughout the analysis, I restrict the sample to individuals from 30 to 65 years who were born between 1945 and 1960. Ameriks and Zeldes (2004) point out that pooling multiple cohorts results in different life-cycle profiles for investment choices depending on whether one controls for time or for cohort effects. Therefore, I focus on individuals born within a relatively short time frame while controlling for age and time effects. In that way, I ensure that all individuals in my sample faced similar environmental conditions at a given age.
I use the waves 1989 until 2016 from the Survey of Consumer Finances (SCF) to measure financial choices of households. The SCF is a triennial repeated cross-section analysis sponsored by the Federal Reserve Board. It is carried out at the household level but collects individual demographic characteristics and income variables as well as detailed information on joint asset holdings of the household. To account for increased likelihood of survey non-response for asset-rich households, the SCF oversamples that population group. Since the focus of my study is not on the very rich, I drop the richest 10% of the sample (in terms of financial wealth). Dropping the upper decile of observations basically affects only couples and hence, does not change the empirical gender investment gap across single men and single

Table 1: Descriptive Statistics: Sample Size

	Individuals	Observations
SCF:		
Single Women	–	2,848
Single Men	–	950
Couples	–	8,513
PSID:		
Single Women	701	10,669
Single Men	593	4,752
Couples	4,050	42,565

women. In contrast, the resulting age profile for average financial wealth of couples is more comparable to measures from other datasets that do not oversample asset-rich households (such as the PSID). Moreover, to ensure the representativeness of the US population, I weigh each observation by the provided survey weight throughout the estimation.

For income variables, labor market outcomes and demographic characteristics I work with data from the Panel Study of Income Dynamics (PSID) spanning from 1989 until 2017. The PSID is a longitudinal panel-survey of private households in the US running from 1968 until today.[2] Besides the core sample, the PSID oversamples low-income families (the 'SEO' sample) and immigrant families (the 'immigrant' sample). To make the sample comparable to that from the SCF, I drop all families belonging to those two sub-samples. Each wave, household members report biographic information, their individual labor force status and individual income levels. All financial variables are converted into 2007 dollars using the CPI-U.[3]

I define a single woman to be a family unit with a female head and no spouse present. Single men are defined accordingly. Table 1 reports the resulting sample size. In total, the PSID sample consists of 57,986 individual-year observations that correspond to 701 single women, 593 single men and 4,050 individuals who live in couples. In contrast, the data drawn from the SCF includes information on 8,513 individuals in couples, on 950 single men and on 2,848 single women.

2.2 Life-Cycle Profiles of Portfolio Allocation

Throughout the analysis, I define financial assets in gross terms, that is financial wealth net of housing assets and debt (i.e. mortgages). Risky assets include direct stock holdings,

[2] Because the Survey of Consumer Finances starts in 1989, I restrict my data sample taken from the PSID to the waves from 1989 until 2017. Data were collected annually until 1997 and afterwards every two years.

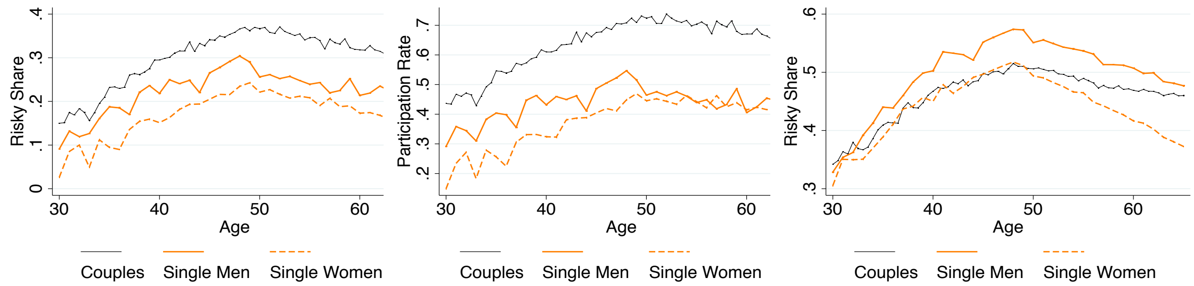
[3] CPI estimates taken from the US Bureau of Labor Statistics, available under this link [Accessed May 22, 2019].

corporate and foreign bonds, the fraction of mutual funds that include the former as well as the fraction of retirement accounts which is invested in stocks.[4]
Figure 1a displays the life-cycle profiles of equity shares for single men, single women and for couples during their working age. Equity shares combine the extensive margin (whether or not the households owns any risky assets) with the intensive margin (conditional on holding risky asset, what portfolio share is allocated to them?). To obtain a more complete picture, Figure 1b and Figure 1c separately plot the stock market participation rate (that is, only the extensive margin) and the conditional risky share (that is, only the intensive margin), respectively. I obtain all graphs by linearly regressing the respective dependent variable on a constant, age, the second polynomial of age, an interaction term of gender and age, a dummy that indicates more than 12 years of education, the number of household members, labor income and year dummies.[5] All differences are statistically different from zero, see Table 2 for the corresponding regression coefficients.

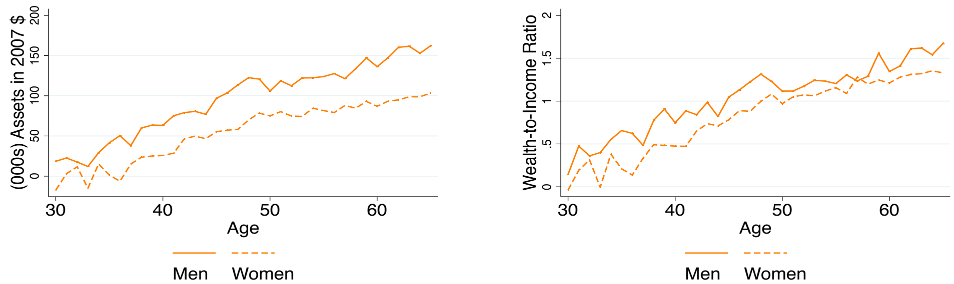
Figure 1a shows that the equity share of single women is lower than that of single men during their entire working life. On average, the equity share of single women is 4.79%pts lower than that of men which – given an average equity share of single men of 23.43% – corresponds to 20.44% and roughly remains constant over the life-cycle. In contrast, the observed gender gap in stock market participation rates (Figure 1b) converges towards the entry to retirement whereas the gender gap in the conditional risky share diverges with age (Figure 1c).
Furthermore, the black solid line in Figure 1a shows that couples have on average a higher equity share than singles which is mainly driven by the extensive margin (see the black solid lines in Figures 1b and 1c, respectively). However, this finding is partly mechanical as couples are composed of two individuals for whom I compute the joint participation probability of participation. If I randomly draw a single men and a single women and compute the likelihood that at least one of them holds risky assets (conditional on age), the participation rate of such a "generated couple" closely aligns with the ones of couples in the data.
Finally, Figure 1d confirms that single women accumulate less wealth than single men what is often referred to as the "gender wealth gap". Over the course of their working life, the gap in financial wealth is on average \$37,760, being largest when entering retirement (\$57,300). The gender wealth gap also prevails when normalizing by current labor income (as shown by the wealth-to-income ratio in Figure 1e).

[4] In Appendix A, I show that my results are robust to adopting a tighter definition of risky assets that does not include retirement accounts.
[5] To account for observations with zero or very little labor income, I transform labor income into its inverse hyperbolic sine before running the regressions.

Figure 1: Life Cycle Pattern of Household Finances (Data)



(a) Equity Share (b) Participation Rate (c) Conditional Risky Share



(d) Financial Assets (Singles) (e) Wealth-to-Income Ratio (Singles)

Notes: Figure 1 plots the life-cycle profiles of the equity share, stock market participation rates and conditional risky shares for singles and couples as well as absolute financial assets and the wealth-to-income ratio of singles. The sample consists of individuals born between 1945 and 1960 in the waves 1989 until 2016 of the Survey of Consumer Finances (SCF). Risky assets are defined as direct stock holdings, corporate and foreign bonds, the fraction of mutual funds that include the former and the fraction of retirement accounts which is invested in stocks.

2.3 Regression Coefficients over the Life-Cycle

The empirical gender differences in portfolio choices reported in Figure 1 can arise due to differences in circumstances or due to differences in preferences. The objective of this paper is to quantify the importance of the former. As a first exercise, I consider linear regressions that control for observable characteristics (Table 2). In particular, I run Tobit regressions (to account for non-participating households) of the equity share on a gender dummy, age polynomials and gender interacted with age (Column (1)). In Column (2), I additionally control for observable characteristics that the literature has shown to be important predictors for portfolio choices.Following Christelis, Georgarakos, and Haliassos (2013), I control for the education of the individual, the overall number of household members and the inverse hyperbolic sine transformation of labor income. Finally, Column (3) furthermore includes the inverse hyperbolic sine transformation of the households' safe financial assets. However, since the amount of safe assets directly affects the equity share (that is defined the amount of risky assets over the sum of safe and risky assets), I treat Column (2) as the main specification throughout the rest of the paper. The corresponding marginal effects for the gender dummy along with their standard errors at various ages are reported in the last three rows of Table

2.[6]

The coefficient indicating whether the individual is a single women is negative (and statistically significant) across all three specifications. In contrast, the interaction term of gender and age is largest in the first column (least controls), slightly smaller in the second column and becomes statistically insignificant in the third specification. When considering the marginal effect of being a single woman on equity shares (see last three rows in Table 2, "ME"), I find a negative and significant gender effect across the entire life-cycle in the third specification that controls for the most observable characteristics. In contrast, in the main specification (Column 2), the "negative" gender effect on equity shares disappears and eventually turns (slightly) positive as individuals age. Thus, the unexplained part of the gender investment gap (i.e. the part that is not accounted for when including controls) is strongest for young households. This remaining part of the gap can either arise because of unobserved heterogeneity across men and women (e.g. in preferences), because the mapping from observable characteristics to portfolio choices is non-linear or because I did not control for the correct observable characteristics.

Therefore, to further explore how differences in circumstances between single men and single women translate into heterogeneous portfolio choices over the life-cycle and to quantify their relative importance, Section 3 builds a structural model of gender and portfolio choice. Having a structural model helps to accommodate non-linearities and to account for factors that affect portfolio choices but cannot be easily controlled for in reduced form specifications, such as expectations and risk exposure.

2.4 On the (Non-)Presence of Housing

The focus of this paper is on liquid financial wealth which is why I abstract from housing wealth both in the empirical part as well as in the model (Section 3). However, in reality, housing constitutes a large share of households' portfolio and housing choices affect stock market behavior.[7] For the purpose of the current analysis, abstracting from housing is a problem if either housing choices directly map into portfolio behavior (and hence, the gender investment gap could be entirely explained by differences in housing) or if it differentially affects portfolio choices by gender, i.e. if housing is an important driver of the gender investment gap itself.

[6] Appendix B lists the corresponding specifications for the participation rate (Table 7) and for the conditional risky share (Table 8).

[7] One of the first papers to introduce housing in a model of portfolio choice were Cocco (2005) and Yao and Zhang (2005). Since then, there has been a large and ongoing literature on housing and portfolio choices, see for example Flavin and Yamashita (2011), Chetty, Sandor, and Szeidl (2017) or Paz-Pardo (2020) to name a few.

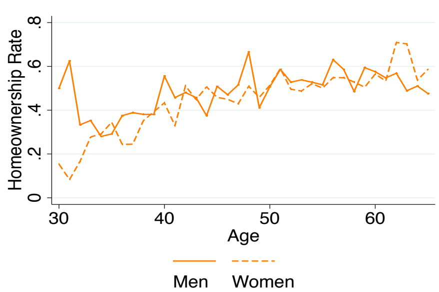
Table 2: Regression Coefficients & Marginal Effects – Equity Shares of Singles

	(1) Equity Share	(2) Equity Share	(3) Equity Share
single woman	-0.367***	-0.150***	-0.0333**
	(0.0153)	(0.0166)	(0.0165)
single woman*age	0.00544***	0.00159***	-0.000253
	(0.000301)	(0.000307)	(0.000267)
age	0.0768**	0.0118	0.00315
	(0.0330)	(0.0440)	(0.0582)
$age^2 * 100$	-0.0538	0.0134	0.0246
	(0.0710)	(0.0940)	(0.125)
$age^3 * 10000$	-0.0302	-0.0312	-0.0411
	(0.0495)	(0.0646)	(0.0858)
High education		0.368***	0.182***
		(0.00560)	(0.00589)
No. of HH members		-0.0511***	-0.0354***
		(0.00293)	(0.00370)
Income		0.0422***	0.0270***
		(0.00118)	(0.00103)
Safe assets			0.0853***
			(0.00147)
Constant	-2.088***	-1.331**	-1.607*
	(0.495)	(0.646)	(0.856)
Observations	4,737	4,735	4,735
Year FE	No	Yes	Yes
ME for women at age 30	-0.0405***	-0.0539***	-0.0485***
	(0.00497)	(0.00380)	(0.00238)
ME for women at mean age (50)	0.1817***	0.0113	-0.0588***
	(0.0160)	(0.0150)	(0.0109)
ME for women at age 65	0.340***	0.0578**	-0.0661***
	(0.0246)	(0.0238)	(0.0186)

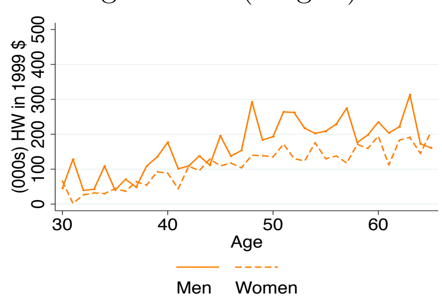
Notes: Estimations are based on Tobit regressions on the sample of individuals that live in households with no spouse present. Source: SCF waves 1989 until 2016; individuals born between 1945 and 1960. Equity Share = Unconditional risky share. *single woman* is a dummy indicating that the household head is a women. *high education* is a dummy equal to one if the household head has more than 12 years of education. *safe assets* refers to safe liquid assets. "ME" indicates the marginal effect of being a women at the respective age. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

To explore whether either of these issues are present in the data, I conduct two exercises: First, if portfolio choices are a direct mapping of housing decisions, we would expect their life-cycle profiles to closely follow those in Figures 1a to 1c. Figure 2 displays the life-cycle profiles of singles for three different housing variables: the homeownership rate, gross housing wealth (henceforth: "HW") and the housing-wealth-to-income ratio (henceforth: "HI"). Single men hold on average slightly more housing wealth than single women. However, I do not find any meaningful gender differences in terms of homeownership rates or in the housing-wealth-to-income ratio despite significant gender differences in portfolio choices both along the extensive and intensive margin. Moreover, the life-cycle patterns for housing variables are different than those of portfolio choices: Neither housing graph displays a hump-shaped pattern with a constant gender gap (as for the equity share), nor a converging gender gap (as for the stock market participation rate) nor a diverging gender gap (as for the conditional risky share).

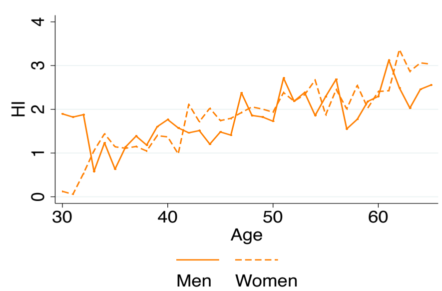
Figure 2: Life Cycle Profiles of Housing Pattern (Singles)



(a) Homeownership Rate



(b) (Gross) Housing Wealth



(c) Housing-Wealth-to-Income Ratio

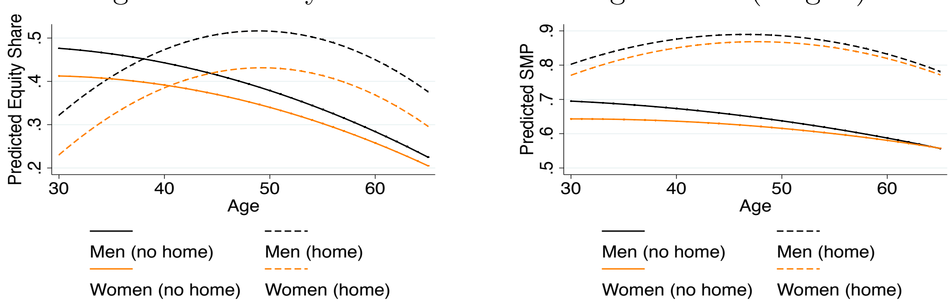
Notes: Figure 2 plots life-cycle profiles of the homeownership rate, gross housing wealth and the housing-wealth-to-income ratio ("HI") for single men and single women. The sample consists of individuals born between 1945 and 1960 in the waves 1989 until 2016 of the Survey of Consumer Finances (SCF).

Second, to understand if housing differently affects the stock market behavior of single men and single women, I compare the predicted equity share and stock market participation rate of single homeowners to those of single non-homeowners, separately by gender. In particular, I first split the sample by housing tenure and run two separate regressions on stock market participation and equity shares, respectively, controlling for observable characteristics. Figure

3 plots the predicted outcome variable from these regressions for an individual with more than 12 years of schooling, the median income of singles, the median number of children for singles (zero) and who is at the respective age in 2001 (which is approx. the midpoint of my sample).

I find that generally, homeownership matters for predicted portfolio choices, in line with previous literature. However, albeit different in levels, the gender differences in predicted participation rates and equity shares are very similar for homeowners and non-homeowners (i.e. the differences between black and the orange line), especially during young age. The predicted gender gap in equity shares is slightly larger among home-owners towards the end of the life-cycle whereas the gap in participation rates and the gap in equity shares for young household does not significantly differ by homeownership-status, increasing my confidence that excluding housing from the analysis does not change the results regarding the sources of the gender investment gap while at the same time keeping the model tractable and easing computational complexity.

Figure 3: Life Cycle Profiles of Housing Pattern (Singles)



(a) Equity Share (b) Participation Rage

Notes: Figure 3 plots the predicted life-cycle profiles of the equity share and the stock market participation rates of a single individual in 2001 who has a high school degree, no children and the medium level of income and safe assets, separately by gender and housing tenure.

3 A Life-Cycle Model of Portfolio Choice

In the following, I develop a life-cycle model of portfolio choice along the lines of Cooper and Zhu (2016). I extend their set-up by introducing three types of households in the model: single men, single women and couples and allow for marital transitions across the life-cycle. In contrast, I abstract from adjustment costs of stock holdings.

3.1 Environment

In the model, agents can be women or men (denote gender by $i = \{f, m\}$) and live either as singles ($\mathcal{S}$) or as a married couple ($\mathcal{M}$). Thus, there exist in total four types of agents:

single women $(\mathcal{S}, f)$, single men $(\mathcal{S}, m)$, married women $(\mathcal{M}, f)$ and married men $(\mathcal{M}, m)$. For all, their life can be split in two stages: working age and retirement. Time is discrete. A model period is one year long. Agents start their life at age 30, retire at age 65 and die deterministically at age 85, i.e. $j \in \{30, 31, ..., 65, ..., 85\}$. They face uncertain survival during retirement that depends on their age j. At age 30, agents are ex-ante heterogeneous in terms of their education θ which can take two values $(\theta = \{l, h\})$ and refers to college and non-college educated individuals in the data. I treat θ as exogenous and assume that agents enter the model after completing education.

During working age, when being single, individuals decide how much to consume (c_i) and how much to save in a safe asset (a_i^s) as well as how much to save in a risky asset (a_i^r). Couples decide jointly on the level of consumption $(c_{\mathcal{M}})$ as well as on how much to save in both types of assets $(a^s_{\mathcal{M}}, a^r_{\mathcal{M}})$. That is, consumption is treated as a public good and becomes private only upon divorce. Moreover, singles face an exogenous marriage probability each period that depends on their gender, age and education level. Likewise, couples face an exogenous divorce probability that again varies by age and both spouses' education.

During retirement, agents do no supply labor but receive a fixed pension which is a fraction of their last realized labor income. Additionally, they face age- and gender dependent medical expenditures and are subject to longevity risk. Upon dying, agents value leaving bequests. As during working age, they can live both as single or couple, however their marital status is fixed (i.e. there is no marriage or divorce). If one spouse living in a couple dies, the surviving spouse continues his or her life as a single with a fraction of the couples' assets to account for increased medical expenditures in the year prior to death as well as for bequests to non-spousal heirs (Jones, De Nardi, French, McGee, and Rodgers, 2020). As before, agents have a portfolio choice between a safe asset and a risky asset.

3.2 Preferences

While I do not allow for gender heterogeneity in preferences, I introduce preference heterogeneity by marital status. In particular, I allow the discount factor β, the coefficient of relative risk aversion γ and the stock market participation costs S^F to vary between singles and couples. Empirically, couples have a higher savings rate than singles albeit their overall income variance being lower, contradicting model predictions. The higher savings rate can arise from various sources, such as saving for children's college, higher homeownership rates among couples or differences in preferences. As the focus of this paper is to explain gender heterogeneity in investment pattern and not differences across singles and couples, I choose to introduce preferences heterogeneity by marital status that allows me to accommodate the data while keeping the model tractable and focusing on the core research question.

3.2.1 Singles

Single individuals can either be a man or a woman with their gender being denoted by $i = \{f, m\}$. They have time-separable CRRA preferences over a consumption good c_i. The period flow of utility is given by:

$$u(c_i) = \frac{\eta_{ij} \left(\frac{c_i}{\eta_{ij}} \right)^{1-\gamma_s}}{1 - \gamma_s}$$

where γ_s is the coefficient of relative risk aversion that is fix across gender and η is an equivalence scale that adjusts for household size and which is allowed to vary by age j and gender i.

3.2.2 Couples

Each couple is composed of exactly one woman and one man. As for singles, couples have time-separable CRRA preferences over the consumption good $c_{\mathcal{M}}$ which is public within the household. Their period flow of utility can therefore be expressed as:

$$u(c_{\mathcal{M}}) = \frac{\eta_{cj} \left(\frac{c_{\mathcal{M}}}{\eta_{cj}} \right)^{1-\gamma_c}}{1 - \gamma_c}$$

Again, γ_c is the coefficient of relative risk aversion (which can be different to that of singles, γ_s) and η is an age-dependent equivalence term adjusting for household size.

3.2.3 Bequest Motive

In the event of death, individuals derive utility from leaving bequests according to:

$$\phi(a') = L \frac{(\omega + a')^{1-\gamma}}{1 - \gamma}$$

where a' denotes the be-quested assets, ω captures the luxuriousness of the bequest motive and L governs the bequest intensity. Bequest preferences are homogeneous across all types of households. Couples value leaving bequests only if they both die within the same period. Whenever only one spouse dies, the surviving spouse continues life as a single with a fraction of the couples' assets (and hence, values leaving bequest in the case of his or her own death).

3.3 Dynamics

3.3.1 Asset Returns

Agent accumulate savings for retirement and to smooth consumption. To do so, they have access to two types of assets: One safe and one risky asset, denoted by a_s and a_r, respectively. The safe asset pays a time-invariant return r_s. In contrast, the return of the risky asset is drawn from the distribution $r_r \sim N(\mu_r, \sigma_r^2)$ that is assumed to be i.i.d and for which it holds that $\mu_r > r_s$. Following Fagereng et al. (2017), I allow for the possibility of stock market crashes and augment the return of the risky asset by a "disaster" state. That is, with probability $(1 - p_{tail})$ the return is drawn from the above normal distribution and with probability p_{tail} a tail event $r_{tail} < \underline{r}_r$ materializes. Short-selling and borrowing are not allowed.

3.3.2 Income Profiles

Following Borella et al. (2019), I assume that income can be split into a deterministic and into a stochastic component. More precisely, income y_{ij} at age j for gender i can be expressed as:

$$y_{ij} = \bar{y}_i \theta_i \xi_{ij} \tilde{y}_{ij}$$

where $\bar{y}_i$ denotes a constant, θ_i is the (exogenous) education premium and ξ_{ij} stands for an age-specific component. Finally, $\tilde{y}_{ij}$ represents the stochastic component of income consisting of a transitory and a persistent shock:

$$\tilde{y}_{ij} = z_{ij} + \epsilon_{\tilde{y}ij}$$

$$z_{i,j+1} = \rho_{zi} z_{ij} + \nu_{zij}$$

where $\epsilon_{\tilde{y}ij}$ and ν_{zij} are independent zero mean random shocks with variances $\sigma^2_{\tilde{y}i}$ and σ^2_{zi} respectively. The parameter $\rho_{zi} \in (0, 1]$ captures the persistence of shock ν_{zi}.
All parameters of the income process are allowed to vary by gender and by marital status to account for the fact that marriage typically results in lower income for women whereas it increases earnings for men.[8] Within couples, the transitory shocks ν_{zfj} and ν_{zmj} are allowed to be correlated as spouses live in the same area and are likely to work in similar industries (family business, they meet at work) and are thus subject to correlated labor market shocks.

[8] The empirical observed "marriage penalty" for women's earnings can arise because of self-selection into marriage (i.e. low income women are more likely to get married) or because marriage itself affects female income, e.g. through childbirth or household specialization. Analyzing the relative importance of these factors is beyond the scope of this paper and therefore, I impose women's earnings drop upon marriage exogenously.

In contrast, following Cocco et al. (2005), labor income shocks are uncorrelated to realizations of the stock return.

3.3.3 Out-of-Pocket Medical Expenditures

When being retired, agents are subject to medical expenditures m_j that are a deterministic function of age and gender. However, because individuals face survival risk and because medical expenditures are strictly increasing in age, deterministic medical expenditures impose a source of risk in the sense that agents are uncertain whether or not they live until a certain age and have to pay the corresponding medical bills. This modeling choice is motivated by De Nardi, French, and Jones (2010) who show that the main sources of risk during retirement are not fluctuations of medical expenditures around its mean but rather their age-dependent level combined with longevity risk.

3.4 Stock Market Participation Cost

In order to avoid the model to predict excess stock market participation rates, agents have to pay a fixed cost S^F each period if they choose to invest part of their savings in the risky asset. This cost is allowed to differ between couples and singles, however, it is equal for single men and single women. As in Vissing-Jorgensen (2002), I model participation costs as a flow variable, that is they have to be paid each period irrespectively of the history of stock holdings. The main advantage to model participation costs as a flow variable rather than an entry cost (see e.g. Alan (2006) or Cooper and Zhu (2016)) is that flow costs do not require introducing stock holdings as an additional state variable and therefore reduce the computational complexity of the model which is – considering the different household structures – already quite substantial.

3.5 Marriage and Divorce

Single individuals get married with an exogenous probability that depends on their gender, their age, their education and on their current productivity realization. Denote this marriage probability by $\mu_i(j, \theta, \tilde{y})$. Conditional on meeting a partner, the probability of meeting a partner with education θ_p and shock realization $\tilde{y}_p$ is:

$$\Pi(.) = \Pi(\theta_p, \tilde{y}_p | \theta_i, \tilde{y}_i)$$

Both partners always have the same age. Individuals are always matched to a partner with the mean empirical amount of assets (conditional on age, gender and education). This specification generates assortative mating along asset holdings as we observe it in the data while at the same time allowing for the possibility that within couples, the income of the husband is usually higher than that of the wife. Couples face an exogenous divorce probability each period that depends on age, the education of each spouse as well as both productivity realizations, that is the likelihood of divorce can be expressed as $\lambda(j, \theta_f, \theta_m, \tilde{y}_f, \tilde{y}_m)$. Upon divorce, assets are split equally between spouses and 25% of assets are destroyed to account for legal fees of divorce and general costs of splitting assets between spouses. There are no alimony payments in order to avoid keeping track of the entire marital history of agents.

3.6 Timing

Timing within one period is as follows. In the beginning of period t all shocks materialize. That is, agents learn their current productivity state(s), their stock market return as well as their marital status. Thus, agents start period t with a given amount of savings that depends on their decisions in period $t-1$, their marital status and the realization of the asset return state. After observing all shock realizations, agents decide on how much to consume and how much to save in both the risky and the safe asset. When investing part of their endowment in the risky asset (i.e. if $a_{r_{t+1}} > 0$), they to pay S^F in the *current* period, that is in period t.

3.7 Recursive Formulation

I express the problem recursively by defining six value functions: the value function for singles, the value function for couples and the value function for an individual living in a couple, all during working age as well as during retirement. The latter is the relevant object when computing the present value of marriage for a single whereas the value function for couples determines the optimal allocation of resources within a couple across time (Borella et al., 2019). Moreover, because the stock market participation cost has to be paid per-period and given the i.i.d nature of the return process for the risky asset, I can combine safe and risky assets into one "asset cash-in-hand" state variable: $a = (1 + r_r)a_r + (1 + r_s)a_s$.

3.7.1 Singles – Working Age

The state variables of a single agent are her gender i, age j, education θ, asset cash-in-hand a and her current income realization $\tilde{y}$. Each period, she has a consumption-savings choice and additionally decides on how to split her savings between the safe and the risky asset.

The corresponding value function reads as:

$$V^S(i,j,\theta,a,\tilde{y}) = \max_{a'_s \geq 0, a'_r \geq 0, c \geq 0} \frac{\eta_{ij}\left(\frac{c}{\eta_{ij}}\right)^{1-\gamma_s}}{1-\gamma_s} + (1-\mu(j,\theta,\tilde{y}))\beta_s s \mathbb{E} V^S(i,j+1,\theta,a',\tilde{y}') +$$
$$\mu(j,\theta,\tilde{y})\beta_s \mathbb{E}\hat{V}^C(i,j+1,\theta,\theta_p,a'+a'_p,\tilde{y}',\tilde{y}'_p)$$

subject to:

$$a'_r + a'_s + c = y(j,\theta,\tilde{y}) + (1+r_s)a_s + (1+r_r)a_r - \mathbb{1}_{a'_r>0} S^F_s$$
$$a = (1+r_r)a_r + (1+r_s)a_s$$

and:

$$\tilde{y} = z + \epsilon_{\tilde{y}} \quad \text{with} \quad z' = \rho_z z + \nu_z \quad \text{and} \quad \epsilon_{\tilde{y}} \sim N(0,\sigma^2_{\tilde{y}}), \nu_z \sim N(0,\sigma^2_z)$$
$$r_r \sim N(\mu_r,\sigma^2_r) \quad \text{with} \quad \mu_r > r_s \quad \text{and} \quad \tilde{y} \perp r_r, \quad \mathbb{E}(\tilde{y}', r'_r, \Pi | j,\theta,z)$$

where η_j denotes an equivalence parameter that controls for changing family size over the life-cycle. $\hat{V}^C$ expresses the value of individual i of getting married to partner p. Single individuals take the expected value over future productivity realizations and asset return when staying single whereas they form expectations over future productivity realization, asset returns and their specific partner in case of getting married.

3.7.2 Singles – Retirement

During retirement, agents do not supply labor and receive a fixed pension income that depends on their last labor income realization. There is no marriage or divorce. Retired individuals face survival risk. Moreover, they are subject to deterministic age-dependent medical expenditures m_{ij}, leaving as state variable gender i, age j, education level θ, asset cash-in-hand a as well as the last income realization before retirement ($\hat{y}$). Each period, retired singles face a consumption-saving as well as a portfolio choice and they value leaving bequests in the case of death.

$$V^S_R(i,j,\theta,a,\hat{y}) = \max_{a'_s \geq 0, a'_r \geq 0, c \geq 0} \frac{\eta_{ij}\left(\frac{c}{\eta_{ij}}\right)^{1-\gamma_s}}{1-\gamma_s} + \beta_s \psi_{ij} \mathbb{E} V^S_R(i,j+1,\theta,a') + \beta_s(1-\psi_{ij}) L \frac{(\omega+a')^{1-\gamma}}{1-\gamma}$$

subject to:

$$a'_r + a'_s + c = pen(\hat{y}) + (1+r_s)a_s + (1+r_r)a_r - m_{ij} - \mathbb{1}_{a'_r>0} S^F_s$$

$$a = (1 + r_r)a_r + (1 + r_s)a_s$$

$$r_r \sim N(\mu_r, \sigma_r^2) \quad \text{with} \quad \mu_r > r_s, \quad \text{and} \quad \mathbb{E}(r_r')$$

where ψ_{ij} denotes the age-dependent survival probability that differs between men and women. Retired singles take the expected value over their next-period asset return as well as their likelihood of survival.

3.7.3 Couples – Working Age

The value function for couples during working age is needed to compute optimal allocation for a couple that consists of a women f and a man m. While I allow for both marriage and divorce during working age, individuals cannot switch partners between two consecutive periods. The state variables of a couple can be summarized by their age j (which is assumed to be the same), education of both spouses θ_f, θ_m, their joint asset holdings a as well as both current productivity realizations $\tilde{y}_f, \tilde{y}_m$. The corresponding value function reads as:

$$\begin{aligned} V^C(j, \theta_f, \theta_m, a, \tilde{y}_f, \tilde{y}_m) = \max_{a_s' \geq 0, a_r' \geq 0, c \geq 0} & \frac{\eta_{\mathcal{M}j} \left(\frac{c}{\eta_{\mathcal{M}j}} \right)^{1-\gamma_c}}{1 - \gamma_c} + \\ & (1 - \lambda(j, \theta_f, \theta_m, \tilde{y}_f, \tilde{y}_m))\beta_c \mathbb{E} V^C(j + 1, \theta_f, \theta_m, a', \tilde{y}_f', \tilde{y}_m') + \\ & \lambda(j, \theta_f, \theta_m, \tilde{y}_f, \tilde{y}_m)\beta_c \sum_{i=f,m} \mathbb{E} V^S(i, j + 1, \theta_i, 0.75\frac{a'}{2}, \tilde{y}_i') \end{aligned}$$

subject to:

$$a_r' + a_s' + c = \sum_{i=f,m} y(j, \theta_i, \tilde{y}_f, \tilde{y}_i) + (1 + r_s)a_s + (1 + r_r)a_r - \mathbb{1}_{a_r' > 0} S_c^F$$

$$a = (1 + r_r)a_r + (1 + r_s)a_s$$

and:

$$\tilde{y}_i = z_i + \epsilon_{\tilde{y}_i} \quad \text{with} \quad z_i' = \rho_{zi} z_i + \nu_{zi} \quad \text{and} \quad \epsilon_{\tilde{y}_i} \sim N(0, \sigma_{\tilde{y}_i}^2), \nu_z \sim N(0, \sigma_{zi}^2) \quad \text{for} \quad i = \{f, m\}$$

$$\begin{pmatrix} \nu_{zf} \\ \nu_{zm} \end{pmatrix} \sim \begin{pmatrix} \sigma_{zf}^2 & \rho_{\sigma_{zf}, \sigma_{zm}} \\ \rho_{\sigma_{zf}, \sigma_{zm}} & \sigma_{zm}^2 \end{pmatrix}$$

$$r_r \sim N(\mu_r, \sigma_r^2) \quad \text{with} \quad \mu_r > r_s, \quad \tilde{y} \perp r_r \quad \text{and} \quad \mathbb{E}(\tilde{y}_f', \tilde{y}_m', r_r' | j, \theta_f, \theta_m, \tilde{y}_f, \tilde{y}_m)$$

Couples take the expected value of both partners' future productivity realizations and joint asset returns when staying married as well as the respective individual's productivity realization and asset return when getting divorced. Moreover, the transitory parts of the income processes (ν_{zf} and ν_{zm}) are allowed to be correlated within couples.

3.7.4 Couples – Retirement

Retired couples receive a flat pension income that depends on the man's last income realization before retirement ($\hat{y}_m$). They do not work and cannot get divorced. However, they individually face the risk of dying. If one spouse dies, the surviving one continues his or her life as single with a fraction δ_i of the couple's assets. If both spouses die within the same period, they jointly value leaving bequests. Their value function reads as:

$$
\begin{aligned}
V_R^C(j, \theta_m, a, \hat{y}_m) = \max_{a'_s \geq 0, a'_r \geq 0, c \geq 0} & \frac{\eta_{\mathcal{M}j} \left(\frac{c}{\eta_{\mathcal{M}j}}\right)^{1-\gamma_c}}{1-\gamma_c} + \beta_c \psi_{jf} \psi_{jm} \mathbb{E} V_R^C(j+1, \theta_m, a', \hat{y}_m) + \\
& \beta_c \sum_{i=f,m} \psi_{ij}(1 - \psi_{-ij}) \mathbb{E} V_R^S(i, j+1, \theta_m, \delta_i a', \hat{y}_m) + \\
& \beta_c (1 - \psi_{jf})(1 - \psi_{jm}) L \frac{(\omega + a')^{1-\gamma}}{1-\gamma}
\end{aligned}
$$

subject to:

$$
a'_r + a'_s + c = pen_c(\hat{y}_m) + (1 + r_s) a_s + (1 + r_r) a_r - \sum_{i=f,m} med_{ij} - \mathbb{1}_{a'_r > 0} S_c^F
$$

$$
r_r \sim N(\mu_r, \sigma_r^2) \quad \text{with} \quad \mu_r > r_s, \quad \text{and} \quad \mathbb{E}(r'_r)
$$

Thus, retired couples take the expected value over their joint asset return as well as the individual's survival probabilities.

3.7.5 Value to an individual of becoming a couple

The value of an individual in a couple is the relevant object when computing the value of single i for getting married to partner p, i.e. the present discounted value of the individual's utility in the event of marriage (Borella et al., 2019). In this context, variables denoted with a $\widehat{hat}$ indicate optimal allocations computed with the value function for couples (section 3.7.3), given the respective state variables. The value of an individual in a retired couple $\hat{V}_R^C$ is defined accordingly.

$$\hat{V}^C(i,j,\theta_i,\theta_p,a,\tilde{y}_i,\tilde{y}_p) = \frac{\eta_j\left(\frac{\hat{c}}{\eta_j}\right)^{1-\gamma_c}}{1-\gamma_c} + (1-\lambda(j,\theta_i,\theta_p,\tilde{y}_i,\tilde{y}_p))\beta_c\mathbb{E}\hat{V}^C(i,j+1,\theta_i,\theta_p,a',\tilde{y}_i',\tilde{y}_p') + \lambda(j,\theta_i,\theta_p,\tilde{y}_i,\tilde{y}_p)\beta_c\mathbb{E}V^S(i,j+1,\theta_i,\frac{a'}{2},\tilde{y}_i')$$

4 Estimation

As in Gourinchas and Parker (2002) or Cagetti (2003), I estimate the model using a two-step strategy. That is, I first estimate all parameters that can be cleanly identified outside of the model and pre-set some parameters to values from the literature. In a second step, I estimate the remaining structural parameters using the Simulated Method of Moments (SMM), taking the parameters from the first stage as given. First stage parameters include initial distributions, parameters related to medical expenditures, the labor income process, survival probabilities and asset returns. I borrow the parameters for the bequest motive (ω, L) from Cooper and Zhu (2016) who estimate bequest parameters in a portfolio choice context with CRRA preferences. Consequently, second stage parameters include the discount factor β, the coefficient of relative risk aversion γ as well as the stock market participation cost S_s^F, all separately for singles and couples. I collect the second stage parameters in the vector $\Theta = \{\beta_s, \beta_c, \gamma_s, \gamma_c, S_s^F, S_c^F\}$.

4.1 First Stage Estimation

4.1.1 Income Profiles

For most individuals, their ability to accumulate wealth and the decision on how to invest that wealth is strongly affected by their life-cycle profile of income. In particular, women typically have lower income than men, leading to heterogeneous financial choices. This gender discrepancy is especially pronounced within couples. Figure 4 shows life-cycle profiles of average income by gender and by marital status. Income is expressed as annual income out of labor earnings (including labor income from farms and businesses), social security benefits and transfers.When estimating those profiles, I restrict the sample to individuals who did not change their education after age 30 because education is exogenous in the model. Moreover, for singles, I include labor earnings, social security benefits and transfers from all members of the households to ensure that my measure of income adequately accounts for disposable income of single households in the data. For couples, I assign each spouse their own labor income, social security benefits and transfers and add half of that from other household

members.[9] Lastly, I drop observations who, according to the described measure, report zero annual income (in the case of couples, if they report zero overall income).
To estimate the income profiles, I follow Borella et al. (2019) and first split the sample by marital status and then separately regress the log of income for individual i at age j,

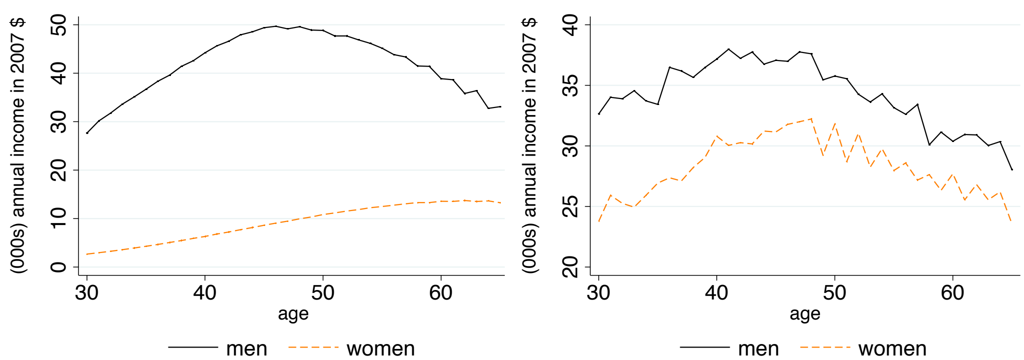
$$ln(income_{ij}) = \alpha + \beta_1 age_{ij} + \beta_2 age_{ij}^2 + \beta_3 woman_i * age_{ij} + \beta_4 woman_i * age_{ij}^2 + \delta_i + u_{ij}$$

on a fixed effect δ_i, age, age^2 as well as their interaction term with a dummy that indicates if the individual is a woman. To obtain shifters for both gender and education level, I regress the sum of the fixed effect and the residual on fully interacted dummies of gender and education level:

$$\delta_i + u_{ij} \equiv w_i j = \gamma_0 + \gamma_1 woman_i + \gamma_2 educ_i + \gamma_2 woman_i * educ_i + \epsilon_{ij}$$

where $educ_i$ is defined as a dummy taking the value one if the respective individuals has more than 12 years of schooling.

Figure 4: Life Cycle Profiles of Income (Deterministic Component)



(a) Couples (b) Singles

The coefficients from these income equations (reported in Table 9 in Appendix D) inform me about the deterministic component of the income process in the model which can be split into a constant, an exogenous education dummy and an age-specific part. I estimate the parameters governing the stochastic component of the income process using the minimum distance estimator as in Guvenen (2009).[10] Table 3 summarizes the results. My point estimates imply a slightly less persistent income process for single women than for single

[9] For some years, the PSID does not separately report transfer income or social security benefits of spouse and household head. In these cases, I allocate half of the overall reported measure to the wife and the other half to the husband.

[10] Details on the estimation strategy for the stochastic part of the income process can be found in Appendix C.

men and than for married individuals. Moreover, the variance of the persistent shock σ_z^2 is higher for single women than for the rest of the population. Thus, single women face overall a more risky income process than single men, however, the difference is relatively small in magnitude. Notably, the income process of married women exhibits a much higher variance of the transitory shock $\sigma_{\tilde{y}}^2$ than that of singles and that of married men.

Table 3: Estimation Results – Stochastic Income Process

Parameter	Men	Women	Men	Women
	Singles		Couples	
ρ_z	0.937	0.9138	0.9307	0.9369
	(0.0131)	(0.0173)	(0.0065)	(0.0048)
σ_z^2	0.087	0.1007	0.0815	0.0882
	(0.021)	(0.0236)	(0.0081)	(0.0097)
$\sigma_{\tilde{y}}^2$	0.1269	0.1175	0.0928	0.2949
	(0.0625)	(0.0417)	(0.0206)	(0.0319)

Notes: Standard Errors in parentheses obtained with bootstrapping (2000 replications).

4.1.2 Marital Transitions

Besides income fluctuations, one main source of financial risk during working age is marital status and possible changes thereof. Figure 5a plots marriage probabilities by age, gender and education whereas Figure 5b displays divorce probabilities by education and age. Both graphs are estimated using PSID data. Marital transitions are defined as the likelihood of getting married (respectively divorced) within the next period conditional on not being married (respectively being married) in the current period. More specifically, I estimate the following logit function, separately for couples and singles:

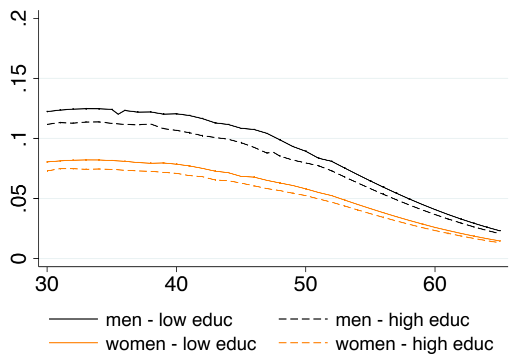
$$\xi_{t+1} = \frac{exp(X_t\beta^s)}{1 + exp(X_t\beta^s)}$$

where ξ_{t+1} denotes the probability of being married (respectively divorced) next period. As explanatory variables, I include the age, age-squared, a dummy indicating whether the individual has some college education as well as a dummy for waves after 1997 to account for switch from annual to biannual frequency in the PSID.[11] Table 10 in Appendix E reports the corresponding regression coefficients.

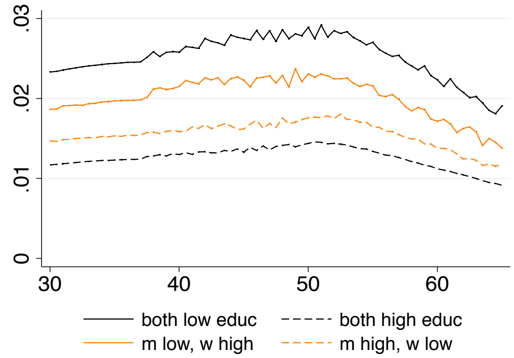
I find that marriage probabilities are higher than divorce probabilities, especially for young individuals. And any given age, single women are less likely than single men to get married

[11] For couples, all demographic variables refer to the household head.

Figure 5: Marital Transition Probabilities



(a) Marriage Probability



(b) Divorce Probability

within the next year. Moreover, the likelihood of divorce displays a hump-shaped pattern whereas the hazard of marriage declines over the life-cycle. The probability of divorce is decreasing in the education of both spouses (Figure 5b). However, couples in which the husband has low education and the wife has a high education are more likely to get divorced than couples whose education is allocated in the opposite way. In contrast, the likelihood of getting married does not significantly differ by education (Figure 5a).

Moreover, I estimate the marriage market (Π) non-parametrically directly from PSID data. Given that marriage occurs exogenously in the model, it may happen that individuals are matched to a partner although they had endogenously chosen to remain single. However, in almost 85% of cases, individuals prefer marriage over singlehood because marriage offers income pooling as well as economies of scale for consumption.

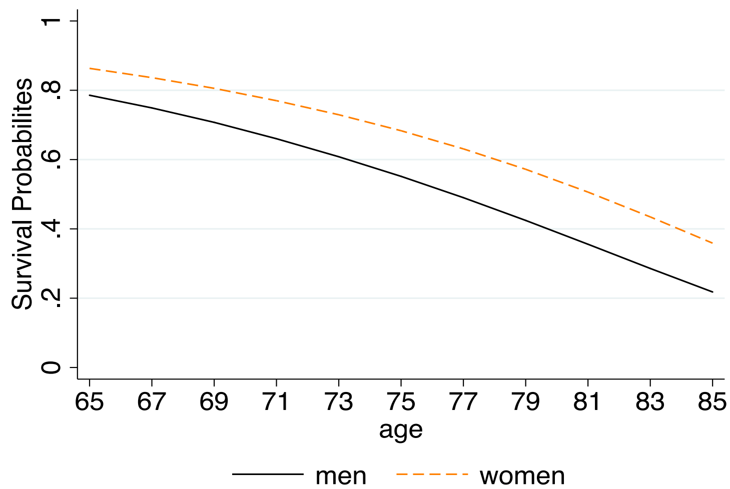
4.1.3 Out-of-Pocket Medical Expenditures

Upon entering retirement, households receive a flat pension income, eliminating their exposure to uninsurable income fluctuations. In contrast, they are subject to out-of-pocket medical expenditures that sharply increase towards the end of the life-cycle. In the model, those expenditures are assumed to be deterministic. However, given that individuals face uncertain survival, medical expenditures impose an uncertainty on households in the sense that it is unclear whether or not the individual survives up to that age. I borrow the parameters describing medical expenditures by age and gender from Borella et al. (2019). The authors estimate deterministic out-of-pocket medical expenditures profiles with data from the HRS separately for men and women who were born in the 1950s. They estimate higher medical expenditures for men at the start of retirement but a steeper gradient for women, especially after age 76. Moreover, to account for the possibility of informal care arrangement among spouses, I assume that medical expenses for married individuals are 80% than that of singles.

4.1.4 Survival Probabilities

Men and women differ in their survival probabilities. Figure 6 shows that, during their retirement phase, women have on average a 12% higher survival probability than men.

Figure 6: Survival Probabilities



Data Source: Life Tables of the US Social Security Population.

I take gender specific death probabilities from the Life Tables of the US Social Security Administration.[12] The death probability at age j is defined as the probability to die within the next year conditional on having survived up to age j. I compute the inverse of those probabilities and work with average values between the years 1990, 2000 and 2010, corresponding to the sample period of my study. In the case of couples, both spouses face individual survival risk and thus, they may die in separate years. If the husband dies, the surviving wife keeps 60% of the household's assets, whereas a surviving husband keeps 70% of the household's asset to account for sharply increasing medical expenses in the year prior to death as well as for bequests to non-spousal heirs (Jones et al., 2020).

4.1.5 Asset Returns

I set the annual return rate of the risk-free asset to 3%. The risky asset has a normally distributed return plus a tail risk. That is, I first assume a risk premium of 3%, and a variance of $Var(\tilde{R}(s)) = \sigma_r^2 = (0.1758)^2$. The latter reflect the variance of the annual total

[12] All tables available under this link [Accessed May 14, 2019].

return index of the S&P 500 from 1989 until 2016. Next, I augment the return of the risky asset by a tail event as in Fagereng et al. (2017). In particular, the return of the risky asset is drawn with probability $1 - p_{tail}$ from the normal distribution and with probability p_{tail} a disaster state materializes. p_{tail} is set to 2% and results in a loss of 50% of all risky assets. A disaster state accounts for severe stock market crashes that we observe in the data but that would not be accounted for when approximating the risky asset return by a normal distribution. Moreover, introducing negative skewness lowers the propensity of agents in the model to invest in risky assets and thus helps to resolve the issue that standard portfolio choice models typically predict excessive equity shares when compared to the data.[13] In the model, the asset return realization is an aggregate shock. When simulating the model for a large set of individuals born between 1945 until 1960 over their life-cycle, I simulate the return of the risky asset to mimic the observed stock market performance in the US when the cohort was in that respective age. In particular, I assume that 20% of the simulated sample were born between 1945 and 1948, 20% between 1949 and 1952 and so on.

4.1.6 Pension Payments

Pension payments are flat and assumed to be 60% of the income during the last year of work. That is, pensions differ by education and productivity state at age 65. Couples receive a common pension which is 1.7 times higher than that of single men.

4.1.7 Equivalence Scales

In the model, the equivalence scales η are allowed to differ by age and household structure (i.e. single men, single women, couple). To compute them, I first estimate the average household size by age and household structure from the PSID and then apply the OECD equivalence scale: I assign a weight of 1 to the first adult household member, a weight of 0.7 to all other adult member and a weight of 0.5 to each child.

4.1.8 Initial Conditions

The initial distribution over asset holdings in the model is chosen such that it mimics the distribution of wealth across individuals born between 1945 and 1960 at age 30 in the SCF. Similarly, I set the fraction of high and low educated individuals by gender to be the average share of individuals with more respectively less than 12 years of schooling in the PSID of

[13] An alternative approach to generate lower equity shares is to introduce adjustment costs for the risky asset (see e.g. Cooper and Zhu (2016)). However, adjustment costs require to introduce the equity share as a state variable. To keep the model tractable, I therefore abstract from these adjustment costs.

that cohort. Finally, the initial distribution of couples and singles is set equal to PSID data for individuals at age 30 born between 1945 and 1960.

4.2 Second Stage Estimation

Taking the parameters from the first stage as given, I estimate the remaining structural parameters $\Theta = \{\beta_c, \beta_s, \gamma_c, \gamma_s, S_c^F, S_s^F\}$ using the Simulated Method of Moments. The exercise is to find $\hat{\Theta}$ that solves the following optimization problem:

$$\mathcal{L} = \min_{\Theta}(M^s(\Theta) - M^d)W(M^s(\Theta) - M^d)' \tag{1}$$

where W represents a weighing matrix, M^d moments derived from the data and $M^s(\Theta)$ their theoretical counterparts derived from model simulations.

4.2.1 Parameter Identification & Choice of Moments

The key challenge is to separately identify the coefficient of relative risk aversion γ, the discount factor β and the stock market participation cost S^F because all parameters directly affect savings behavior and portfolio choices of households. Hence, different parameter values are not entirely orthogonal to one another which makes their separate identification difficult. In this section, I provide (informal) intuition why my moments of choice are informative about the parameters in question. Once households cross the threshold of participation, the participation cost S^F becomes irrelevant for their decision on how much to invest in the risky asset. Taking this discrepancy into account, I identify γ by exploiting heterogeneity in the portfolio share across *participating* households, that is, in the conditional risky share. Moreover, I use heterogeneity in wealth levels to identify β and S^F to match the life-cycle profiles of participation rates. In particular, I target the life-cycle profiles for the conditional risky share, for the participation rate and for absolute wealth levels of couples to identify parameters referring to couples in the model $(\beta_c, \gamma_c, S_c^F)$. For parameters that govern the choices of singles $(\beta_s, \gamma_s, S_s^F)$, I target the corresponding life-cycle profiles of single men. Consequently, life-cycle profiles of single women serve as untargeted moments to validate the model.

4.2.2 The Weighting Matrix W

I first estimate the 2nd stage parameters by using the identity matrix, i.e. with $W = \mathcal{I}$. Consequently, every moment receives equal weight in the estimation procedure. In a second run, I use the inverse of the variances of my moment conditions as a (diagonal) weighting matrix in order to assign a lower weight to less precisely estimated data moments ($W = \frac{1}{\mathcal{V}}$).

This approach follows Cooper and Zhu (2016) and is in contrast to most papers that use the standard variance-covariance matrix (e.g. Cagetti (2003) or Alan (2006)). However, in the current set-up, different moments are based on different sample sizes: While the participation rates and wealth levels include all observations, the conditional risky share only includes stock market participants. Hence, I could only estimate covariances for the restricted sample of stockholders which is not necessarily more informative than the diagonal matrix.

5 Quantitative Results

5.1 2nd Stage Parameters

Table 4 reports the estimated second stage parameters. I take the bequest parameters from Cooper and Zhu (2016) which results in $L = 0.128$ and $\delta = 0.73$. My results imply that singles discount the future more (larger β) and display a slightly higher risk aversion (larger γ) than couples. In contrast, the estimated stock market participation cost is considerably larger for couples than it is for singles.

In the case that uses the inverse variance matrix to weigh its moments, the estimate for the per period stock market participation cost corresponds to an annual cost of \$615.5 for couples and of \$326.5 for singles.[14] The coefficient of relative risk aversion is in contrast more similar: While my estimates suggest $\gamma_s = 2.677$ for singles, I find $\gamma_c = 2.654$ for couples. These values (especially for the coefficient of risk aversion) are at the lower end of estimates introduced by previous papers of portfolio allocation that augment a normally distributed return of the risky asset by a tail event. Generally, portfolio choice models have difficulties in matching participation rates and equity shares without introducing risk preferences or stock market participation costs that seem unlikely high when compared to life-cycle models without a portfolio choice. Introducing negative skewness in the return of the risky asset helps to address this puzzle. Moreover, in the current set-up, marriage and divorce introduce a dimension of financial risk for agents that so far has been largely overlooked in the household finance literature. Therefore, my model is able to match equity shares of households with relatively low degrees of risk aversion and standard values for the participation costs. Fagereng et al. (2017) estimate an annual stock market participation cost of \$69 but also introduce quite a high degree of risk aversion with $\gamma = 11$. In contrast, Catherine (2019) estimates a CRRA coefficient of $\gamma = 8.2$ and an annual stock market participation cost of $\$1,010$.[15] The estimates for β are in line with previous literature.

[14] As this is work in progress, I focus for now on the case that uses the inverse variance matrix only.

[15] In his paper, Catherine (2019) addresses the trade-off that life-cycle models of portfolio choice either require a very high degree of risk aversion (typically in combination with a very low discount factor) or a very high stock market participation cost to match the data by introducing cyclical skewness in labor earnings.

Cooper and Zhu (2016) estimate a discount factor of 0.869, Fagereng et al. (2017) of 0.77 and Catherine (2019) of 0.92. However, given that my coefficients for the relative risk aversion are well below all of their estimates, my estimates for β are comparably low. One reason is that I exclude housing wealth and target the life-cycle profile of financial wealth instead of net worth. Moreover, the possibility of divorce increases the precautionary savings motives for couples while at the same time generating high-asset single households (who got divorced) that are absent in models with only bachelor households.

Table 4: Estimated 2nd Stage Parameters

β_c	β_s	γ_c	γ_s	S_c^F	S_s^F	$\mathcal{L}$	W
0.81	0.881	2.52	2.551	\$600	\$325	20874.76	$\mathcal{I}$
0.7911	0.8793	2.654	2.677	\$615.5	\$326.5	459.9713	$\frac{1}{\mathcal{V}}$

5.2 Model Fit

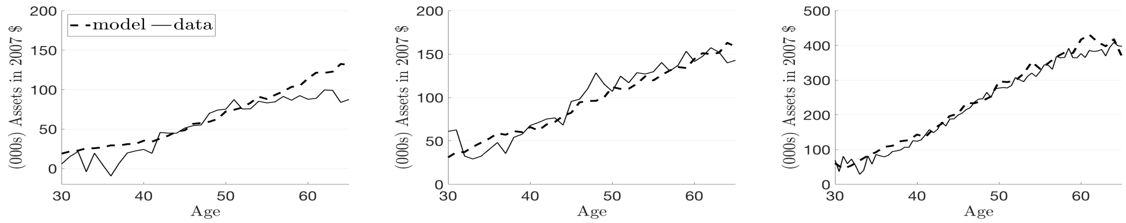
Figures 7 and 8 contrast the life-cycle profiles of equity shares and asset accumulation from the data (see section 2) with those generated by the model. The model is able to replicate the targeted evolution of wealth (Figure 7) for couples and single men. Moreover, it matches very well the un-targeted asset accumulation of single women over their life-cycle. Figure 8 illustrates the fit for the equity share: While the fit is quite good for couple, the model over-predicts the equity share of single men early in the life-cycle. However, most importantly, the model is able to capture the gap in equity shares between single men and single women (and hence, matches the life-cycle profile of equity shares for single women) without introducing preference heterogeneity by gender while at the same time matching overall asset accumulation by household structure.

When splitting the equity share along the extensive margin (participation rate) and the intensive margin (conditional risky share), I find that the model does a good job at matching the life-cycle profiles of participation rates, especially for single households (Figure 9). Finally, Figure 10 shows the model fit for life-cycle profiles of the conditional risky share. In contrast to the data, the model predicts the conditional risky share to be declining in age. This difficulty of portfolio choice models to match the life-cycle profiles of conditional risky shares is common: Because labor income is uncorrelated to the asset return, it acts as a substitute

To make his results comparable to mine, the listed values refer to the case when he estimates his model without cyclical skewness.

for the safe asset. Therefore, a decreasing human-to-financial wealth ratio over the life-cycle translates into a declining optimal risky share as individuals age. Nevertheless, the model correctly matches the average levels of conditional risky shares for singles.

Figure 7: Model Fit of Asset Accumulation



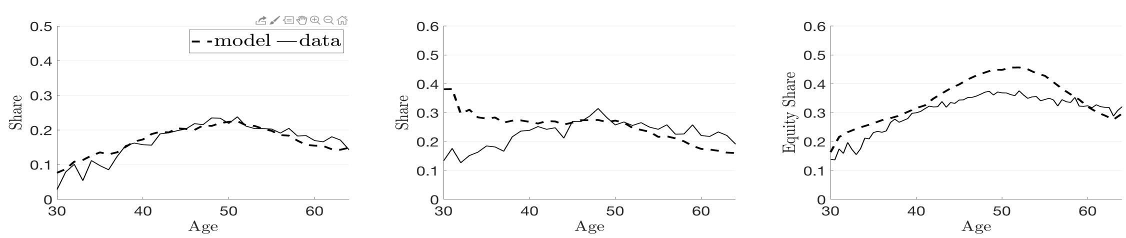
(a) Single Women (b) Single Men (c) Couples

Notes: Figure 7 plots the model fit of asset accumulation for single women, single men and couples. The solid lines show the data (as plotted in Figure 1) whereas the dashed line display the simulated life-cycle profiles generated from the model.

5.3 Simulated Regressions

To compare the reduced-form results from Table 2 with those generated by the model, Table 5 replicates the same regression on simulated data. In particular, Column (1) shows the regressions estimates from model simulations whereas Column (2) re-reports the regression results from the main empirical specification (compare Table 2, Column (2)). All of these coefficients are un-targeted in the estimation exercise. The model slightly over-predicts the gender investment gap, especially early in life and hence, the coefficient for "single women" is more negative on the simulated data than it is on empirical data. In contrast, the interaction term of being a single woman and age is is more positive in the simulated dataset, resulting in an under-prediction of the investment gap as individuals age. When comparing the marginal effects, I find that the simulated data captures well the declining gender investment gap over the life-cycle. In particular, in both specifications, reduced form regressions that control for

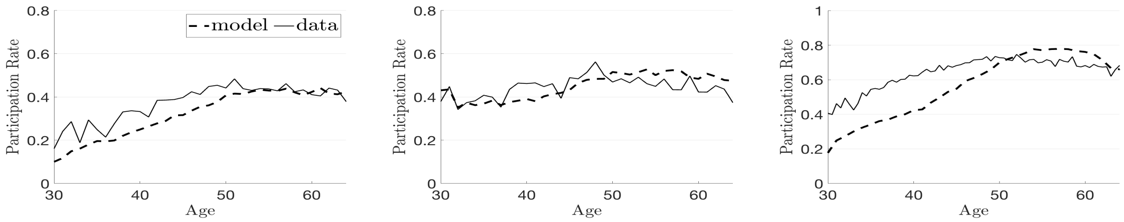
Figure 8: Model Fit of Equity Shares



(a) Single Women (b) Single Men (c) Couples

Notes: Figure 8 plots the model fit of equity shares for single women, single men and couples. The solid lines show the data (as plotted in Figure 1) whereas the dashed line display the simulated life-cycle profiles generated from the model.

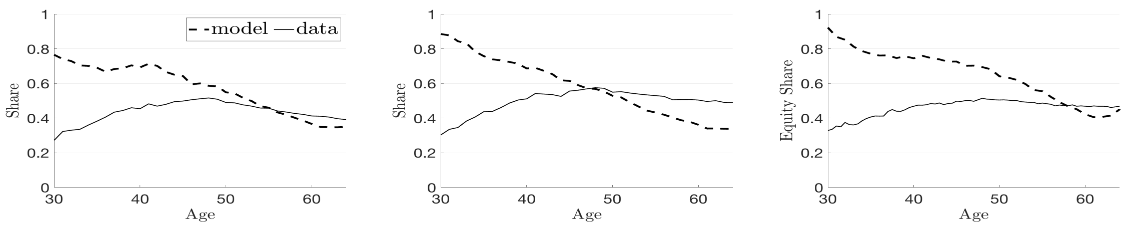
Figure 9: Model Fit of Participation Rates



(a) Single Women (b) Single Men (c) Couples

Notes: Figure 9 plots the model fit of participation rates for single women, single men and couples. The solid lines show the data (as plotted in Figure 1) whereas the dashed line display the simulated life-cycle profiles generated from the model.

Figure 10: Model Fit of Conditional Risky Shares



(a) Single Women (b) Single Men (c) Couples

Notes: Figure 10 plots the model fit of conditional risky shares for single women, single men and couples. The solid lines show the data (as plotted in Figure 1) whereas the dashed line display the simulated life-cycle profiles generated from the model.

observable characteristics can explain gender differences in portfolio choices for individuals beyond age 45 but fail to fully explain the gap for younger households. Thus, the marginal effect for gender early in the life-cycle in reduced form regressions remains statistically significant if the underlying data generating process assumes preferences homogeneity across men and women. Hence, it appears that either factors which cannot as easily be controlled for (such risk exposure, expectations) or non-linearities explain the residual part of the gap. To uncover these factors to quantify their relative importance over the life-cycle, Section 6 performs several counterfactual exercises.

6 Counterfactual Simulations

6.1 Decomposing the Gender Investment Gap

In this Section, I decompose the gender gap in equity shares and in wealth levels along the dimensions of gender heterogeneity within the model, that is along income levels, income risk (productivity), marital transition probabilities, the expected characteristics of the partner in the event of marriage (the "marriage market": Π), the distribution across education levels, initial wealth levels at age 30, differences in household size (which is captured by the equiv-

Table 5: Regression Coefficients & Marginal Effects – Equity Shares of Singles

Equity Share	(1) Model Simulations	(2) Data SCF
single woman	-0.319***	-0.150***
	(0.0798)	(0.0166)
single woman*age	0.00363**	0.00159***
	(0.00157)	(0.000307)
age	-0.146**	0.0118
	(0.0646)	(0.0440)
$age^2 * 100$	0.383***	0.0134
	(0.138)	(0.0940)
$age^3 * 10000$	-0.298***	-0.0312
	(0.0965)	(0.0646)
High education	-0.0124	0.368***
	(0.0167)	(0.00560)
No. of HH members		-0.0511***
		(0.00293)
Income	0.472***	0.0422***
	(0.0112)	(0.00118)
Constant	-3.500***	-1.331**
	(0.984)	(0.646)
Observations	4,737	4,735
Year FE	No	Yes
ME for women at age 30	-0.101***	-0.0539***
	(0.0222)	(0.00380)
ME for women at mean age	0.0328	0.0113
	(0.0753)	(0.0150)
ME for women at age 65	0.154	0.0578**
	(0.127)	(0.0238)

Notes: Estimations are based on Tobit regressions on the sample of individuals that live in households with no spouse present. Column (1) are model simulations, column (2) refers to data from the SCF waves 1989 until 2016; individuals born between 1945 and 1960. Equity Share = Unconditional risky share. *single woman* is a dummy indicating that the household head is a women. *high education* is a dummy equal to one if the household head has more than 12 years of education. *safe assets* refers to safe liquid assets. "ME" indicates the marginal effect of being a women at the respective age. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

alence scale η) as well as medical expenses and age-dependent survival probabilities during retirement. In particular, I replace the female value for each channel with that of men and study the resulting gender gaps in asset holdings and in equity shares. Table 6 shows the results. The column "Model" reports the gender investment gap in the respective counterfactual scenario whereas the column "% explained" indicates how much of the baseline gap can be explained through the respective channel.

In general, aggregate portfolio allocations in the model are determined by the policy function for the optimal risky share $\alpha = \phi(X)$, conditional on state variables X, and the distribution of individuals across the state space. Thus, differences in investment behavior between the baseline model and the counterfactual scenario can arise because the distribution of individuals across the state space changes (*"composition effect"*) or because individual decision rules at any given point in the state space differ (*"policy effect"*).

Decomposing the Gap in Wealth Levels

Table 6 shows that differences in income levels, income risk and in household size explain the largest fraction of the wealth gap between single men and single women. Lower income levels naturally translate into less asset holdings, explaining 22.92% of the "gender wealth gap". At the same time, the income process of single women is less risky than that of single men. Therefore, assigning single women the male income risk increases female precautionary savings, reducing the gender gap in asset holdings. This channel in isolation explains on average 6.49% of the gap. Moreover, larger household sizes of single women act as a consumption commitment and lower the ability to save. On average, differences in household size between single men and single women explain 31.55% of the gender gap in wealth levels. Furthermore, giving single women the male marriage probabilities (that is, increasing their marriage hazard conditional on age) decreases the wealth gap by 9.41%. Especially during young age, agents in the model prefer marriage over being single because couples can pool their income and enjoy economies of scale for consumption. Increasing the likelihood of such a positive financial outcome consequently reduces precautionary savings.

The remaining channels are quantitatively less important for explaining gender heterogeneity in asset holdings. Assigning women the male medical expenses, the male survival probability or the male partner's characteristics in the event of marriage (marriage market) lowers asset holdings of single women and consequently increases the gender wealth gap. In the counterfactual scenario, lower medical expenses in very old age (beyond age 76) combined with a smaller chance of surviving up to that point decrease the incentive of single women to accumulate precautionary savings against longevity risk. However, as most gender differences in survival risk and in medical expenditures materialize at the very end of the life-cycle, this effect is quantitatively small when averaging over the working life.

Finally, when simulating the model under the assumption that both single men and single women start from the same (male) wealth level at age 30 substantially reduces the wealth gap early in life, increasing consumption of young single women.

Decomposing the Gap in Equity Shares

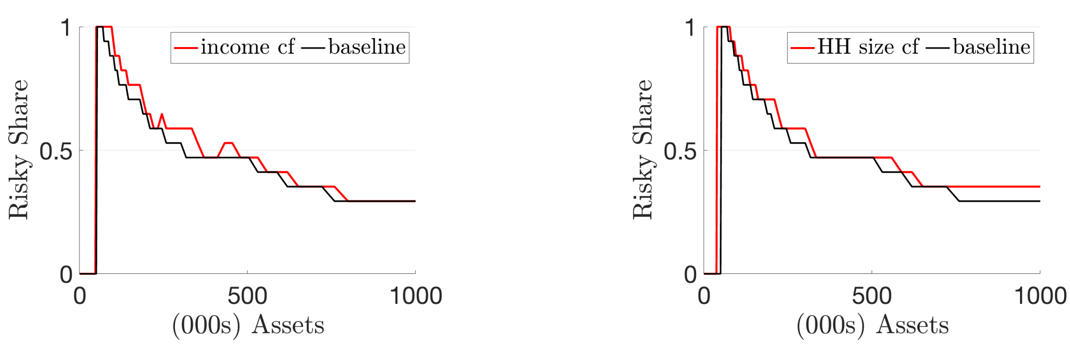
After having examined the gender gap in wealth levels, the next step is to decompose the gap in equity shares, that is the "gender investment gap". Naturally, as portfolio allocations are closely related to asset holdings, differences in both income levels and in household size not only explain the largest share of the gender gap in wealth levels but also in equity shares. When single women receive the male income level, the simulated sample is composed of richer individuals who are more likely to cross the participation threshold of risky asset holdings. Figure 11a plots the policy function for the risky share from the baseline model (black line) and from the income counterfactual (red line). It shows that in addition to this *composition effect*, single women are also more willing to invest in the risky asset conditional on their wealth level (and other states). Because of the bond-like nature of labor income, a higher human capital endowment (i.e. more expected income in future periods) increases the willingness of single women to invest in the risky asset for a given level of wealth and current income. Thus, the effect of changing income levels on the gender investment gap operates both through the policy effect as well as through the composition effect. This result illustrates why controlling for current income in reduced-form regressions is not sufficient to explain the overall effect of income on portfolio choices. In addition to increased wealth through higher income today (and in previous periods), current portfolio choices are also affected by *expectations* over future income realization. Hence, it is rational for a single woman who has the same current income, wealth and education level as a single man to invest less risky because she expects a lower income in the future.

The same mechanism applies when lowering female household sizes to that of single men. The sample composition changes because smaller household sizes decrease per-period consumption and consequently translate into higher wealth levels. Moreover, the decision rules for a given point in the state space become more risky (see Figure 11b). Assigning single women the male household size not only decreases their consumption needs today but also in future periods, making them less vulnerable to financial risk and thus increasing their willingness to invest in the risky asset. Quantitatively, assigning single women the male income level reduces the gender gap in equity shares by 39.11%, whereas eliminating heterogeneity in household sizes narrows the gap by 41.16%.

In contrast, when single women face the same income risk as single men, the gender gap in equity share widens by 1.35%. The income process for single men is more volatile than for single women (see Table 3). Hence, giving single women the male income process lowers their willingness to invest in the risky asset. This "negative" effect on the gender gap in equity shares prevails despite the simulated sample being composed of on average richer single women (who are more likely cross the participation threshold) because of increased precautionary savings.

Moreover, if single women had the same marriage probability as single men (that is, increasing the likelihood of marriage), the gap in equity shares widens by 32.37%. The effect of increased marriage probability on equity shares mainly occurs because women hold less precautionary savings in the counterfactual scenario and are thus less likely to cross the threshold of risky asset participation. Moreover, especially early in life, it its optimal for single women to allocate a larger share of their portfolio towards the risky asset (conditional on participating) because they expect a higher household income (through marriage) in future periods. Generally, the significance of expected marriage for equity shares highlights the importance of considering marriage and divorce as a substantial financial risk when explaining the life-cycle behavior of portfolio choices. In particular, because marriage probabilities strongly differ by gender, they are key in explaining investment differences between single men and single women.

Figure 11: Policy Function for Risky Share (Single Women)



(a) Income counterfactual

(b) HH size counterfactual

Notes: Figure 11a plots the policy function for the risky share against current assets. The black line shows the policy function of the baseline model whereas the red line displays the policy function for the counterfactual scenario in which I assign single women the deterministic part of income process of single men ("income counterfactual") or the male household sizes ("HH size counterfactual"). The policy functions refer to single women with high education and medium-high labor productivity at age 40.

Table 6: Decomposition Results

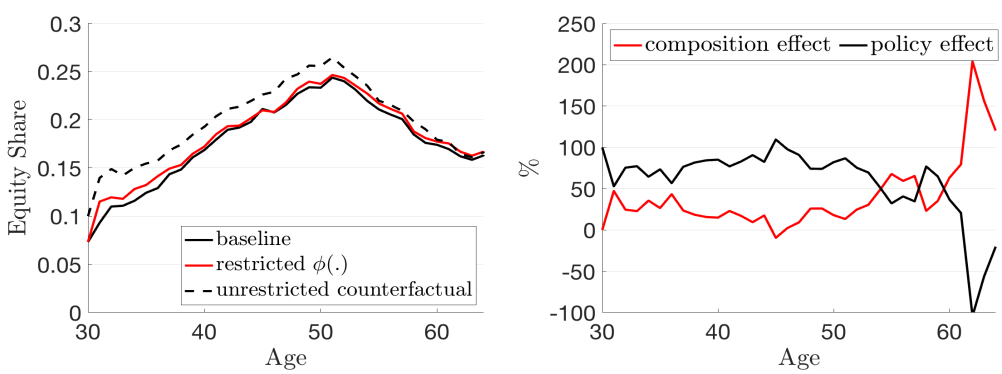
Gap in (000s) Asset Holdings in 2007 $	Model	% explained	Data
Baseline	31.89		43.52
Male income level	24.58	22.92%	
Male income risk	29.82	6.49%	
Male HH size	21.83	31.55%	
Male marriage probability	34.89	-9.41%	
Male marriage market	33.33	-4.52%	
Male education distribution	32.78	-2.79%	
Male medical expenses	32.35	-1.44%	
Male survival probability	32.57	-2.13%	
Male initial wealth	26.04	18.36%	
Gap in Equity Share	**Model**	**% explained**	**Data**
Baseline	5.19%		5.93%
Male income level	3.16%	39.11%	
Male income risk	5.26%	-1.35%	
Male HH size	2.95%	43.16%	
Male marriage probability	6.87%	-32.37%	
Male marriage market	5.39%	-3.85%	
Male education distribution	5.47%	-5.39%	
Male medical expenses	5.05%	2.7%	
Male survival probability	5.45%	-5.0%	
Male initial wealth	4.18%	18.82%	

6.2 Composition vs. Policy Effect

Assigning single women the male income level increases the aggregate equity share of single women not only because their distribution across the spate space changes but also because it is optimal for them to invest more risky, conditional on state variables (see Figure 11a). Whereas I did control for the former when running reduced form regressions on empirical and simulated data (Table 5), it is not as straightforward to include expectations about future income levels in these regressions. The objective of this section is to quantify the relative importance of expectations versus current income differences on the gender investment gap along the life-cycle. To do so, I simulate the income counterfactual and restrict the policy functions of the risky share to be the same as in the baseline model. Consequently, any difference in life-cycle profiles between this simulation and the counterfactual with unrestricted policy functions can be attributed to the policy effect, that is, because single women choose a more risky portfolio allocation conditional on their state vector (through different expectations about future income). Figure 12 plots the results of this exercise. Figure 12a contrasts the life-cycle profiles of the baseline model (black solid line), the counterfactual with unrestricted policy functions (black dashed line) and the counterfactual in which I restrict the policy function for the risky share to be the same as in the baseline (red line). Any difference between the red line and the black dashed line can be attributed to the policy effect, thus, to changes in portfolio choices arising from different expectations. To quantify the relative importance of the policy effect vs. the composition effect, Figure 12b plots what percentage of the change in female equity shares in the income counterfactual can be explained by the composition effect (red line) and how much by the policy effect (black line).

I find that the policy effect almost entirely explains differences in equity shares between the baseline model and the income counterfactual early in the life-cycle. That is, larger equity shares for relatively young individuals arise from differences in decision rules (because of larger human capital endowments) rather than from differences in the sample composition. However, as individuals age, the composition effect becomes more important, eventually overtaking the policy effect at around age 56. Over the life-cycle, the remaining human capital endowment decreases and hence, its impact on policy functions becomes smaller. In contrast, higher income levels in previous years have translated into more savings, affecting the sample composition. In line with this finding, reduced form regressions (see Table 5) can explain the gender investment gap later in life whereas they fail do to so for younger households.

Figure 12: Composition vs. Policy Effect of Income Counterfactual (Single Women)



(a) Life-Cycle Profiles (b) Composition vs. Policy

Notes: Figure 12 decomposes the difference in female equity shares between the baseline model and the income counterfactual into a composition and into a policy effect. Figure 12a contrasts the life-cycle profiles from the baseline model (black solid line) to the unrestricted income counterfactual (black dashed line) and to the counterfactual that restricts the policy function for the risky share to be the same as in the baseline model (red line). In Figure 12b, the composition effect (red line) shows which percentage can be explained through on average richer individuals in the simulated sample whereas the policy effect (black line) shows which percentage can be explained by differences in decision rules for equity shares, fixing all other state variables; Both lines mechanically add up to 100 at every age.

7 Conclusion

This paper studies the determinants of the gender investment gap over the life-cycle. It first provides empirical evidence that women allocate a smaller share of their liquid portfolio into risky assets while at the same time being less likely to hold any risky assets at all. Reduced form regressions reveal that the gender investment gap remains statistically significant after controlling for observable characteristics such as household size or income level, especially for young households. In contrast, an estimated structural portfolio choice model that restricts preferences to be equal across gender but allows for heterogeneity in observable characteristics and stochastic processes is able to (over-)explain the gap. Counterfactual simulations reveal that higher income levels of single men account for 39.11% in the observed gender gap in equity shares whereas gender differences in household sizes of singles explain 43.16%. Most importantly, the structural analysis finds that both contemporaneous income levels and household sizes as well as the *expected path* of these variables matter for current-period investment behavior. Because of the bond-like nature of labor income, a higher human capital endowment increases an agent's optimal equity share for any given level of wealth. Similarly, lower expected household sizes reduce future consumption needs and increase financial risk-taking. The effect of future realizations on portfolio choices is stronger for young households and hence, reduced form regressions that do not take into account households' expectations have troubles explaining the gender investment gap early in life.

References

Agnew, J., P. Balduzzi, and A. Sunden (2003): "Portfolio choice and trading in a large 401 (k) plan," *American Economic Review*, 93, 193–215.

Alan, S. (2006): "Entry costs and stock market participation over the life cycle," *Review of Economic Dynamics*, 9, 588–611.

Almenberg, J. and A. Dreber (2015): "Gender, stock market participation and financial literacy," *Economics Letters*, 137, 140–142.

Ameriks, J. and S. P. Zeldes (2004): "How do household portfolio shares vary with age," Tech. rep., working paper, Columbia University.

Andersen, S., J. Marx, K. M. Nielsen, and L. Vesterlund (2020): "Gender Differences in Negotiation: Evidence from Real Estate Transactions," *NBER WP No 27318.*

Barber, B. M. and T. Odean (2001): "Boys will be boys: Gender, overconfidence, and common stock investment," *The Quarterly Journal of Economics*, 116, 261–292.

Bertocchi, G., M. Brunetti, and C. Torricelli (2011): "Marriage and other risky assets: a portfolio approach," *Journal of Banking & Finance*, 35, 2902–2915.

Bogan, V. L. and J. M. Fernandez (2017): "How Children with Mental Disabilities Affect Household Investment Decisions," *American Economic Review*, 107, 536–40.

Borella, M., M. De Nardi, and F. Yang (2019): "The lost ones: the opportunities and outcomes of non-college educated Americans born in the 1960s," *NBER WP No 25661.*

Cagetti, M. (2003): "Wealth accumulation over the life cycle and precautionary savings," *Journal of Business & Economic Statistics*, 21, 339–353.

Catherine, S. (2019): "Countercyclical Labor Income Risk and Portfolio Choices over the Life-Cycle," *HEC Paris Research Paper No. FIN-2016-1147.*

Chetty, R., L. Sandor, and A. Szeidl (2017): "The effect of housing on portfolio choice," *The Journal of Finance*, 72, 1171–1212.

Christelis, D., D. Georgarakos, and M. Haliassos (2013): "Differences in Portfolios across Countries: Economic Environment versus Household Characteristics," *The Review of Economics and Statistics*, 95, 220–236.

Christiansen, C., J. S. Joensen, and J. Rangvid (2015): "Understanding the effects of marriage and divorce on financial investments: the role of background risk sharing," *Economic Inquiry*, 53, 431–447.

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

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

Cook, D. O., R. Kieschnick, and B. D. McCullough (2008): "Regression analysis of proportions in finance with self selection," *Journal of empirical finance*, 15, 860–867.

Cooper, R. and G. Zhu (2016): "Household finance over the life-cycle: What does education contribute?" *Review of Economic Dynamics*, 20, 63–89.

Croson, R. and U. Gneezy (2009): "Gender differences in preferences," *Journal of Economic Literature*, 47, 448–74.

Cubeddu, L. and J.-V. Ríos-Rull (2003): "Families as Shocks," *Journal of the European Economic Association*, 1, 671–682.

De Nardi, M., E. French, and J. B. Jones (2010): "Why do the elderly save? The role of medical expenses," *Journal of Political Economy*, 118, 39–75.

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

Flavin, M. and T. Yamashita (2011): "Owner-occupied housing: life-cycle implications for the household portfolio," *American Economic Review*, 101, 609–14.

Goldsmith-Pinkham, P. and K. Shue (2020): "The Gender Gap in Housing Returns," *Mimeo*.

Gomes, F. and A. Michaelides (2005): "Optimal life-cycle asset allocation: Understanding the empirical evidence," *The Journal of Finance*, 60, 869–904.

Gourinchas, P.-O. and J. A. Parker (2002): "Consumption over the life cycle," *Econometrica*, 70, 47–89.

Guvenen, F. (2009): "An empirical investigation of labor income processes," *Review of Economic Dynamics*, 12, 58–79.

Hubener, A., R. Maurer, and O. S. Mitchell (2015): "How family status and social security claiming options shape optimal life cycle portfolios," *The Review of Financial Studies*, 29, 937–978.

Jones, J. B., M. De Nardi, E. French, R. McGee, and R. Rodgers (2020): "Medical Spending, Bequests, and Asset Dynamics Around the Time of Death," *NBER WP No 26879*.

Ke, D. (2018): "Cross-Country Differences in Household Stock Market Participation: The Role of Gender Norms," in *AEA Papers and Proceedings*, vol. 108, 159–62.

Love, D. A. (2010): "The effects of marital status and children on savings and portfolio choice," *The Review of Financial Studies*, 23, 385–432.

Lusardi, A. and O. S. Mitchell (2014): "The economic importance of financial literacy: Theory and evidence," *Journal of Economic Literature*, 52, 5–44.

Merton, R. C. (1971): "Optimum consumption and portfolio rules in a continuous-time model," *Journal of Economic Theory*, 3, 373–413.

Paz-Pardo, G. (2020): "Homeownership and Portfolio Choice over the Generations," .

Säve-Söderbergh, J. (2012): "Self-Directed Pensions: Gender, Risk, and Portfolio Choices," *The Scandinavian Journal of Economics*, 114, 705–728.

Sunden, A. E. and B. J. Surette (1998): "Gender differences in the allocation of assets in retirement savings plans," *The American Economic Review*, 88, 207–211.

Thörnqvist, T. and A. Olafsson (2019): "Bargaining over risk: The impact of decision power on household portfolios," .

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

Vissing-Jorgensen, A. (2002): "Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures," *NBER WP No 8884*.

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

A Supplementary Figures

One concern in my study could be that the gender investment gap mainly arises through asset holdings in retirement accounts. If single men are more likely to hold retirement accounts (e.g. because of their job types or employment histories) than single women, and if individuals, regardless of gender, tend to invest retirement savings more risky than other types of wealth, the gender investment gap would reflect gender heterogeneity in the labor market rather than in investment choices. Therefore, Figure 13 plots the life-cycle profiles of equity shares, stock market participation rates and conditional risky shares for singles and couples based on a tighter definition of risky assets that excludes savings held through retirement accounts. It shows that, compared to the baseline, the gender gap in equity shares actually increases (Figure 13a), alleviating concerns that empirical investment differences across single men and single women are mainly driven through savings that are linked to certain types of jobs.

Figure 13: Life Cycle Pattern of Household Finances – Excluding Retirement Accounts

(a) Equity Share (b) Participation Rate (c) Conditional Risky Share

Notes: Figure 13 plots the life-cycle profiles of the equity share, stock market participation rates and conditional risky shares for singles and couples. The sample consists of individuals born between 1945 and 1960 in the waves 1989 until 2016 of the Survey of Consumer Finances (SCF). Risky assets are defined as direct stock holdings, corporate and foreign bonds as well as the fraction of mutual funds that include the former. In contrast to Figure 1, financial assets do not include wealth held through retirement accounts.

B Regression Coefficients and Marginal Effects

Table 7: Regression Coefficients & Marginal Effects – Participation Rates of Singles

	(1) SMP	(2) SMP	(3) SMP
single women	-0.278***	-0.130***	-0.0377**
	(0.0161)	(0.0158)	(0.0159)
single woman * age	0.00470***	0.00206***	0.000588*
	(0.000309)	(0.000291)	(0.000305)
age	-0.0287**	-0.0413**	-0.0898***
	(0.0139)	(0.0184)	(0.0222)
$age^2 * 100$	0.114***	0.106***	0.204***
	(0.0308)	(0.0387)	(0.0477)
$age^3 * 10000$	-0.112***	-0.0841***	-0.153***
	(0.0222)	(0.0266)	(0.0332)
High education		0.269***	0.114***
		(0.00402)	(0.00487)
No. of HH members		-0.0305***	-0.0187***
		(0.00197)	(0.00159)
Income		0.0290***	0.0175***
		(0.000552)	(0.000547)
Safe assets			0.0559***
			(0.000593)
Constant	0.464**	0.415	0.974***
	(0.201)	(0.273)	(0.328)
Observations	4,737	4,735	4,735
R^2	0.014	0.126	0.293
Year FE	No	Yes	Yes
ME for women at age 30	0.00452	-0.00639***	-0.00244
	(0.00338)	(0.00257)	(0.00319)
ME for women at mean age (50)	0.197***	0 .0777***	0.0216
	(0.0154)	(0.0138)	(0.0151)
ME for women at age 65	0.334***	0.138***	0.0387
	(0.0243)	(0.0222)	(0.024)

Notes: Estimations are based on OLS on the sample of individuals that live in households with no spouse present. Source: SCF waves 1989 until 2016; individuals born between 1945 and 1960. SMP = Stock Market Participation. *single woman* is a dummy indicating that the household head is a women. *high education* is a dummy equal to one if the household head has more than 12 years of education. *safe assets* refers to safe liquid assets. "ME" indicates the marginal effect of being a women at the respective age. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 8: Regression Coefficients & Marginal Effects – Conditional Risky Share of Singles

	(1) Cond. Share	(2) Cond. Share	(3) Cond. Share
single women	0.0515	0.0620*	0.0177
	(0.0320)	(0.0332)	(0.0252)
single woman * age	-0.00234***	-0.00245***	-0.00175***
	(0.000628)	(0.000647)	(0.000511)
age	0.172***	0.0923	0.0499
	(0.0593)	(0.0602)	(0.0571)
$age^2 * 100$	-0.302**	-0.169	-0.0740
	(0.124)	(0.128)	(0.120)
$age^3 * 10000$	0.170**	0.101	0.0374
	(0.0850)	(0.0880)	(0.0828)
No. of HH members		-0.00901**	-0.0176***
		(0.00371)	(0.00345)
Income		0.00304***	0.00543***
		(0.000841)	(0.000754)
Safe assets			-0.0604***
			(0.000671)
Constant	-2.635***	-1.244	-0.0939
	(0.918)	(0.919)	(0.880)
Observations	2,173	2,173	2,173
R^2	0.034	0.054	0.223
Year FE	No	Yes	Yes
ME for women at age 30	-0.0887***	-0.085***	-0.0871***
	(0.00649)	(0.00632)	(0.00672)
ME for women at mean age (50)	-0.187	-0.188***	-0.161***
	(0.0324)	(0.0331)	(0.0275)
ME for women at age 65	-0.252***	-0.256***	-0.209***
	(0.0498)	(0.0511)	(0.0417)

Notes: Estimations are based on OLS on the sample of individuals that live in households with no spouse present. Source: SCF waves 1989 until 2016; individuals born between 1945 and 1960. Cond. Share = Risky Share conditional on Participation. *single woman* is a dummy indicating that the household head is a women. *high education* is a dummy equal to one if the household head has more than 12 years of education. *safe assets* refers to safe liquid assets. "ME" indicates the marginal effect of being a women at the respective age. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

C Income Process Estimation – Stochastic Component

I estimate the stochastic component of the income process by the minimum distance estimator as in Guvenen (2009). In particular, I assume the unexplainable part of the income process (that is, the residual term ϵ_{it} from the income equation) to follow a persistent-transitory process:

$$\tilde{y}_j = z_j + \epsilon_{\tilde{y}}$$

$$z_{j+1} = \rho_z z_j + \nu_z$$

A persistent-transitory process requires identification of three parameters: The persistence parameter ρ_z, the variance of the persistent shock $\sigma_{\epsilon^2_{\tilde{y}}}$ and the variance of the transitory shock $\sigma^2_{\nu_z}$ which can be identified by the following moments:

$$\frac{cov(\tilde{y}_j, \tilde{y}_{j-2})}{cov(\tilde{y}_{j-1}, \tilde{y}_{j-2})} = \frac{\rho_z^2 var(z_{j-2})}{\rho_z var(z_{j-2})} = \rho_z$$

$$var(\tilde{y}_{j-1}) - \frac{cov(\tilde{y}_j, \tilde{y}_{j-1})}{\rho_z} = var(z_{j-1}) + \sigma_{\epsilon_{\tilde{y}}} - var(z_{j-1}) = \sigma_{\epsilon_{\tilde{y}}}$$

$$var(\tilde{y}_{j-1}) - cov(\tilde{y}_j, \tilde{y}_{j-2}) - \sigma_{\epsilon_{\tilde{y}}} = \rho_z^2 var(z_{j-2}) + \sigma_{\nu_z} + \sigma_{\epsilon_{\tilde{y}}} - \rho_z^2 var(z_{j-2}) - \sigma_{\epsilon_{\tilde{y}}} = \sigma_{\nu_z}$$

I recover the parameters that minimize the distance between the covariance-variance matrices of the income process in the data and their theoretical counterparts under the assumption that $Var(z_{-1}) = 0$. However, the PSID only collects data every two years after 1997 while the model is written in annual frequency. To account for this inconsistency, I linearly interpolate income for individuals that I observe in two consecutive waves for the missing year in which no PSID data was collected. I run four different estimations for married men, married women, single men and single women. Table 3 in the main text displays the results.

D Income Process – Deterministic Component

Table 9: Regression Coefficients for Income Estimation (Deterministic Component)

	(1) First Stage	(2) Second Stage	(3) First Stage	(4) Second Stage
	Couples		Singles	
high educ.		0.660***		0.618***
		(0.0248)		(0.0398)
Woman		-3.432***		-0.573***
		(0.0476)		(0.0428)
Woman*high educ.		-0.484***		0.204***
		(0.0593)		(0.0527)
age	0.148***		0.0553**	
	(0.0251)		(0.0224)	
$age^2 * 100$	-0.155***		-0.0672***	
	(0.0260)		(0.0225)	
age*woman	0.0449***		0.00580	
	(0.00582)		(0.00539)	
Constant	6.144***	1.390***	9.841***	-0.140***
	(0.610)	(0.0203)	(0.557)	(0.0320)
Observations	34,280	34,280	5,722	5,722
Number of unique indiv.	3,284		892	
R^2	0.017	0.348	0.008	0.169

Notes: Estimations are based on (fixed-effect) OLS regressions from PSID Data, waves 1989-2017 on individuals born between 1945 and 1960. Corresponding Figure is Figure 4 in the main text. Dependent variable of first stage: Log of annual income (labor income, social security income and transfers) of the household head. In years where social security (transfer) income is not available separately by head and spouse, I use combined social security (transfer) income and assign it 50-50 to both spouses. For singles, I add labor income, social security benefits and transfers from other household members. For couples, I again split the income from other household members 50-50 between spouses. Dependent variable of second stage: fixed effects plus residual from first stage. *high educ.* is a dummy equal to one if the individual has more than 12 years of schooling; *Woman* is a dummy indicating if the individual is woman; Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

E Marriage and Divorce Probabilities

Table 10: Regression Coefficients for Marriage and Divorce Hazards

	(1) Marriage Prob.	(2) Divorce Prob.
Woman	-0.466***	
	(0.102)	
Age	0.134**	0.0705
	(0.0611)	(0.0513)
$age^2 * 100$	-0.201***	-0.0897
	(0.0675)	(0.0548)
$\mathbb{1} > 1997$	0.218	0.325***
	(0.135)	(0.109)
High Educ. (Head)	-0.112	-0.476***
	(0.101)	(0.0922)
High Educ. (Spouse)		-0.221**
		(0.0903)
Constant	-4.190***	-5.060***
	(1.363)	(1.179)
Observations	7,489	38,104

Notes: Estimations are based on Logit regressions from PSID Data, waves 1989-2017 on individuals born between 1945 and 1960. Corresponding Figure is Figure 5 in the main text. Dependent variable: Likelihood of getting married (resp. divorced) within the next year, conditional on not being married (resp. being married) today. The age of a couple is the average age of both spouses. For education within couple, head refers to the husband and spouse refers to the wife. In contrast, singles are always labeled as head. *High Educ.* is a dummy equal to one if the individual has more than 12 years of schooling; *Woman* is a dummy indicating if the individual is woman; $\mathbb{1} > 1997$ indicates observations that were interviewd after 1997 to account for the changing frequency of the PSID. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

F Wealth Distribution

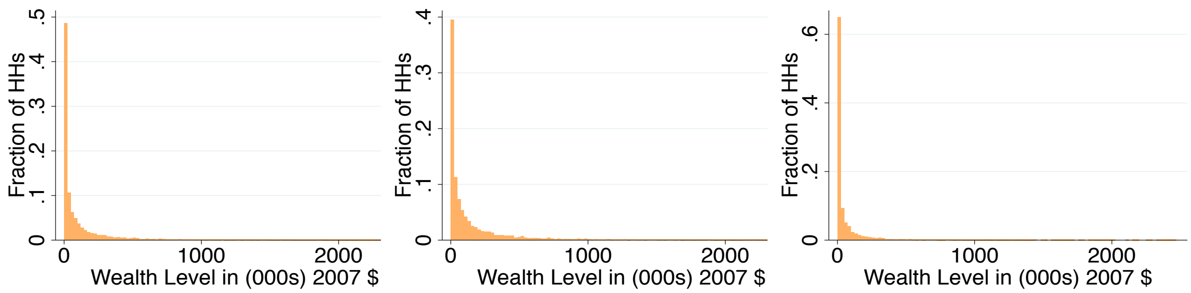
In this section, I compare additional (non-targeted) wealth statistics from the data of individuals who report to hold between \$0 and \$2,5 Mio. in financial assets with those simulated by the model. First, Table 11 shows the fraction of wealth held by quintiles of the entire sample, couples and singles. Second, Figure 14 and 15 display histograms of the wealth distribution in the data and from model simulations, respectively. The model endogenously replicates a right-skewed wealth distribution, meaning that the majority of individuals hold zero or little assets, whereas very few households hold large fractions of overall wealth. However, the generated wealth distribution is still less skewed than in the data. Moreover, I find that empirically, singles' wealth is more dispersed than that of couples. In contrast, the model generates similar wealth distributions across marital states.

Table 11: Wealth Distribution: Data vs. Model (non-targeted)

Quintile	Data	Model			
Whole Sample:					
1	0.037%	0.73%			
2	0.75%	3.65%			
3	4.01%	9.8%			
4	14.20%	22.3%			
5	81.01%	63.51%			
Quintile	Data	Model	Quintile	Data	Model
Couples:			*Singles:*		
1	0.12%	0.81%	1	0.006%	0.83%
2	1.40%	3.97%	2	0.27%	4.38%
3	5.44%	10.6%	3	2.30%	10.79%
4	16.32%	23.16%	4	11.01%	23.43%
5	76.73%	61.45%	5	86.41%	60.57%

Notes: Table 11 compares the wealth distribution in the data with model simulation. Each quintile indicates how much of overall wealth is owned by the respective quintile. Data refers to the cohort of individuals born between 1945-1960 taken from waves 1989-2016 of the Survey of Consumer Finances (SCF). Wealth refers to financial wealth as specified in the main text.

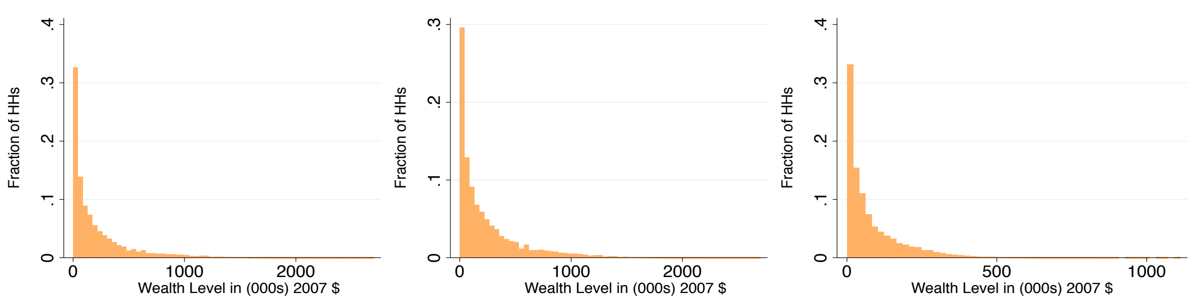
Figure 14: Wealth Distribution – Data



(a) Whole Sample (b) Couples (c) Singles

Notes: Figure 14 plots the wealth distribution of the whole sample, couples and singles from the data. The sample consists of individuals born between 1945 and 1960 in the waves 1989 until 2016 of the Survey of Consumer Finances (SCF). Wealth refers to (gross) financial wealth.

Figure 15: Wealth Distribution – Model (non-targeted)



(a) Whole Sample (b) Couples (c) Singles

Notes: Figure 15 plots the wealth distribution of the whole sample, couples and singles from simulating 100,000 households over their life-cycle with the model described in Section 3.