Genetic Endowments, Income Dynamics, and Wealth Accumulation Over the Life-Cycle*

Daniel Barth[1], Nicholas W. Papageorge[2,3,4], Kevin Thom[5], and Mateo Velásquez-Giraldo[2]

[1]Federal Reserve Board of Governors
[2]Johns Hopkins University
[3]IZA
[4]NBER
[5]University of Wisconsin - Milwaukee

Abstract: A growing literature on gene-by-environment ($G \times E$) interactions — much of it concerned with human capital — asks how policy and other environmental factors affect inequality related to genetic endowments. Though typically absent from this literature, economic models could play a vital role in understanding such interactions. We estimate a life-cycle model of consumption, savings, and portfolio decisions, allowing genetic endowments linked to education to directly affect income, labor disutility, stock market participation costs, and asset returns. Even accounting for completed education, childhood SES, and inheritances, we estimate that these genetic factors increase wealth both through life-cycle income profiles and through rates of return on invested wealth. Counterfactual exercises predict how social security reforms would modify these relationships. Strikingly, some policies may simultaneously flatten gene-wealth gradients but increase gene-welfare inequality. Economic theory has a valuable role to play in revealing such distinctions in the analysis of $G \times E$ interactions.

Keywords: Wealth, Inequality, Genetic Endowments, Saving and Portfolio Choices, Retirement.
JEL Classification: D14, D15, D31, D63, G50, H31, H55, I38, J24, J26.

*First draft: July 1, 2022. Current draft: August 16, 2023. We are grateful for helpful comments from Christopher Carroll, Michael Keane, and Jonathan Skinner, along with seminar participants at the University of St. Andrews, the Board of Governors at the Federal Reserve, Wisconsin - Madison, Wisconsin - Milwaukee, Banco de la República de Colombia, Yale, SITE, and NBER-Aging 2022, along with two conferences on genes and economic analysis: Genes, Social Mobility, and Inequalities across the Life-Course and Frontiers in Economic Analysis with Genetic Data. We also acknowledge excellent research assistance from Matthew Gonzalez. Kevin Thom acknowledges funding from the National Institute on Aging of the National Institutes of Health (R01AG055654 and R56AG058726).Views and opinions expressed are those of the authors and do not necessarily represent official positions or policy of the Federal Reserve Board of Governors.

†dannybarth@gmail.com, papageorge@jhu.edu, kthom.work@gmail.com, mvelasq2@jhu.edu.

1. Introduction

Decades of research have shown that genetic variation influences complex behaviors and outcomes, including educational attainment, income, and wealth. One strand of this literature exploits data on twins and adoptees to decompose variance in such outcomes into genetic and non-genetic components and consistently finds a non-trivial role for genetic factors (Taubman, 1976; Plug and Vijverberg, 2003; Cesarini et al., 2009; Cronqvist and Siegel, 2014, 2015; Black et al., 2020; Fagereng, Mogstad, and Rønning, 2021). More recently, breakthroughs in behavioral genetics have led to the increasing availability of molecular genetic variables in rich survey data. As a result, researchers can directly estimate associations between specific genetic variants (often aggregated into summary measures called *polygenic indices* or *polygenic scores*) and economic outcomes (Rietveld et al., 2013; Okbay et al., 2016; Lee et al., 2018; Hill et al., 2019). Much of this literature has studied polygenic indices constructed to predict educational attainment, as these indices have proven remarkably powerful in predicting a series of important economic outcomes over the life-cycle, including not only educational attainment, but also income and household wealth.[1]

Although there are many sources of inequality, genetic factors are distinct because they are fixed at conception and automatically transmitted across generations. As long as genetic factors remain a black box (an inscrutable "nature" contrasted with a controllable "nurture"), it may be tempting to assume that there is little room to counter resulting inequality with policy. In contrast, demystifying the primitive abilities and preferences linking genetic differences to complex outcomes like income or wealth can help to clarify the role of public policy. Indeed, the existing literature emphasizes the study of how genetic and environmental factors interact to shape complex economic outcomes (gene-by-environment or $G \times E$ interactions). Such research often estimates whether specific reforms or other government actions moderate or amplify genetic associations and thus offers guidance about how policy can mitigate genetic inequality. However, nearly without exception, $G \times E$ studies approach the problem atheoretically, without an optimizing model of individual behavior.[2] We ar-

[1]See Beauchamp et al. (2011); Benjamin et al. (2012); Visscher et al. (2017) for reviews.

[2]A lone exception is offered by Biroli (2015), who estimates an optimizing model of health behaviors with heterogeneity in genetic predisposition to obesity. In our current work, we examine a substantially different set of behaviors, emphasize the possibility of $G \times E$ interactions in welfare, and explore discrepancies between these interactions and the $G \times E$ interactions in observable outcomes

gue that this approach, while it highlights a role for policy, severely limits what can be learned from $G \times E$ findings. In particular, incorporating genetic variation into economic models is important to understand whether $G \times E$ findings generalize to counterfactual policies and whether these findings also point to $G \times E$ interactions in welfare.

We advance the study of genetic endowments and economic outcomes by developing and estimating a model of consumption, savings, and portfolio choice to understand the channels through which genes predicting education affect wealth accumulation over the life-cycle. Motivated by descriptive empirical patterns in Barth, Papageorge, and Thom (2020) and Papageorge and Thom (2020), the model allows genetic variation (measured by a polygenic index for educational attainment) to affect wealth accumulation through multiple channels beyond completed education, including labor income, utility from work, and financial proficiency. We explicitly model childhood SES, later-life inheritances, and their correlations with genetic endowments to address confounding with family environments. A realistic representation of pensions and the Social Security system allows us to simulate the effects of cost-saving reforms that have been proposed as responses to an aging population. We estimate model parameters using data on polygenic indices, income, retirement decisions, wealth, and portfolio choices in the Health and Retirement Study.

The first contribution of this analysis is to quantify several different channels beyond completed schooling through which genetic endowments previously linked with education contribute to wealth inequality. The most important channels are differences in labor income and financial sophistication as measured by stock market returns. Differences in stock market returns emerge as particularly important. Our estimates suggest that, on average, a one standard-deviation increase in the polygenic score for educational attainment is associated with expected log-returns from risky investments that are 78 basis points (0.78 percentage points) higher per year. This gradient induces a positive correlation between the polygenic score and stock market participation, as individuals forgo wealth-building stock market investments not only due to costs of entry but also due to lower expected returns (Calvet, Campbell, and

emphasized by the existing literature. While Houmark, Ronda, and Rosholm (2020) estimate the parameters of a production function for childhood skills, they do not estimate the kind of optimizing model needed to conduct welfare analyses or explore counterfactual behavioral responses.

Sodini, 2007). These results highlight how the genetic endowments studied here have a compounding effect over the life-cycle—the same genetic factors that give individuals an advantage in acquiring educational attainment also increase income, and these advantages are jointly magnified by higher portfolio returns over the life-cycle.

A second contribution is to use the estimated model to conduct an *ex-ante* analysis of $G \times E$ interactions. Given that the structural estimates highlight the importance of portfolio returns in driving the gene-wealth gradient, it is natural to ask whether this gradient can be affected by the design of public pension systems, since these are a critical policy tool used to address inequality in wealth and income at older ages. Thus, we examine the extent to which two counterfactual Social Security reforms would affect the strength of the polygenic score–wealth gradient in our estimated model. The first increases the age of retirement by delaying the schedule of available benefits. The second leaves the schedule in place but reduces the level of benefits. These policies exemplify two of the main types of reforms that have been considered to address fiscal challenges of an aging population (Social Security Administration, 2022b). We find that the policies differ in the margins along which they incentivize individuals to adjust: the shift in the schedule of benefits pushes agents to extend their working lives, and the reduction in benefits incentivizes them to save more. Strikingly, while the benefit reduction policy reduces *wealth inequality* between individuals with different genetic endowments, it increases *welfare inequality* since it induces reductions in consumption among a group of individuals whose consumption is already low. This finding underscores the importance of not only considering $G \times E$ interactions in observable outcomes but also in terms of the welfare of individuals since these may differ and generate conflicting policy prescriptions. Naively, one might assume that the empirical association between genes and wealth is a good proxy for the relationship between genes and welfare. Under such an interpretation, any policy that reduces gene-wealth gradients could be seen as reducing inequality in welfare generated by the genetic lottery. Our analyses demonstrate that this interpretation can be perilous and that economic theory can be valuable for interpreting the welfare implications of the expanding $G \times E$ literature.

Additional counterfactual exercises reveal the important role that differences in financial sophistication play in determining the relative welfare consequences of each

Social Security reform. Even though the median welfare effects of the two counterfactual reforms are similar under our baseline estimates, the welfare costs of the benefit-reduction plan fall substantially in alternative scenarios where individuals' financial inefficiencies are smaller. Thus, the benefit-reduction policy may be more attractive than the shift in the schedule of benefits if the financial proficiency of the general population can be improved through other policies.

Our analysis connects to four main literatures. First, our paper is related to the literature that studies genetic associations and $G \times E$ interactions using polygenic scores, surveyed by Biroli et al. (2022). A series of influential genome-wide association studies (GWAS) have estimated associations between individual genetic markers and complex socioeconomic outcomes, allowing for the construction of polygenic scores for these traits (Rietveld et al., 2013; Okbay et al., 2016; Lee et al., 2018; Karlsson Linnér et al., 2019; Hill et al., 2019). Polygenic scores for educational attainment have received particular attention, with several papers documenting associations between such scores and completed education, income, wealth, and other measures of socioeconomic success (Belsky et al., 2016, 2018; Barth, Papageorge, and Thom, 2020; Papageorge and Thom, 2020). Rustichini et al. (2023) estimate a model of intergenerational transmission that separates the influence of a polygenic score for education into channels running through cognitive and non-cognitive skills. Past studies have also found that relationships between these scores and educational attainment are moderated by environmental factors including compulsory schooling laws, school quality, birth order, and family socioeconomic status (Barcellos, Carvalho, and Turley, 2021; Arold, Hufe, and Stoeckli, 2022; Trejo and Domingue, 2018; Muslimova et al., 2020; Ronda et al., 2022; Papageorge and Thom, 2020; Belsky et al., 2016). Other studies examine interactions between environmental factors and polygenic scores for other outcomes including alcohol use, smoking, heart disease, and depressive symptoms (Fletcher and Lu, 2021; Bierut et al., 2023; Baker et al., 2022; Furuya et al., 2022). We contribute to this literature by incorporating genetic variation captured by a polygenic score into a structural life-cycle model with multiple mechanisms linking genes and wealth. This allows us to quantify these channels and perform *ex-ante* counterfactuals to assess the likely $G \times E$ effects of policies that have been proposed but not implemented. Importantly, our analysis highlights that the

$G \times E$ interactions in outcomes uncovered by this literature might be different than $G \times E$ interactions in welfare and that economic theory is useful in distinguishing between these cases.

Our paper also engages the literatures on the heritability of economic outcomes, and causes of intergenerational persistence in socioeconomic status. A large literature uses twin and adoption studies to demonstrate that genetic variation has explanatory power for outcomes like earnings, risk-taking and giving, investment decisions and biases, and saving decisions (Taubman, 1976; Cesarini et al., 2009, 2010; Cronqvist and Siegel, 2014, 2015; Black et al., 2020; Fagereng, Mogstad, and Rønning, 2021). Our analysis is not well suited to answer questions about the overall extent of heritability for the human capital and wealth outcomes that we discuss. In general, polygenic scores tend to explain only a fraction of the variance that these other methods attribute to genetic factors (Becker et al., 2021). Rather, our goal is to take advantage of the fact that polygenic scores are observed in rich longitudinal data sets like the HRS to more easily analyze the mechanisms through which they operate and how they interact with environments. Such findings may be particularly important in the literature that models intergenerational mobility and tries to account for sources of intergenerational persistence (Gayle, Golan, and Soytas, 2022; Rustichini et al., 2023; Collado, Ortuño-Ortín, and Stuhler, 2023). Such models inevitably have to make choices about what parameters are heterogeneous and how such heterogeneity is transmitted across generations. Our results offer guidance to such modelling efforts. In particular, they suggest that it may be important to allow genetic endowments related to educational attainment to also affect household income and wealth dynamics and to properly model how such channels could affect the intergenerational transmission of human capital.

A third literature uses life-cycle models to study the fiscal and welfare consequences of policies related to retirement. Wealth at retirement varies substantially, and life-cycle models can provide benchmarks and tests to assess the adequacy of households' savings (Hubbard, Skinner, and Zeldes, 1995; Bernheim, Skinner, and Weinberg, 2001; Scholz, Seshadri, and Khitatrakun, 2006). Heterogeneity across dimensions like earning potential, life expectancy, and household structure is a crucial characteristic of models used to evaluate reforms to the Social Security system (Conesa and Krueger, 1999; Fuster, İmrohoroğlu, and İmrohoroğlu, 2007; Hairault, Langot, and Sopraseuth, 2008; Imrohoroğlu and Kitao, 2012). Indeed, the degree to which a household benefits or loses from a particular reform depends on its characteristics (Fuster, İmrohoroğlu, and İmrohoroğlu, 2003; Kaygusuz, 2015; Bagchi, 2019). Our model has a rich representation of heterogeneity along observable and unobservable dimensions, with more than five thousand types of *ex-ante* differentiated agents. We add to the literature by incorporating a novel measure of genetic endowments and allowing it to influence financial sophistication, earnings, and the cost of work. Having a common driver of these characteristics generates compounding welfare effects from policy changes: the people who are most reliant on Social Security payments are those least well-positioned to offset changes using their private savings or extending their working lives. For each of the alternative policies that we analyze, we examine the distribution of expected welfare losses across the population and identify those who are most harmed.

Lastly, our paper relates to the literature on household portfolio choice and the wealth distribution. Differences in financial sophistication can generate large differences in wealth across households over time (Lusardi, Michaud, and Mitchell, 2017). Investigating the sources of such differences and how they affect choices is crucial to understanding wealth inequality. The workhorse models in the portfolio choice literature prescribe that every household should allocate a substantial fraction of their wealth to stocks (Merton, 1969; Samuelson, 1969; Cocco, Gomes, and Maenhout, 2005). Given the generally low rates of stockholding documented in various countries, later studies have incorporated financial costs to stockholding, heterogeneity in preferences, and tail-risks as ways to reconcile model predictions with household decisions (see e.g., Vissing-Jorgensen, 2002; Gomes and Michaelides, 2005; Fagereng, Gottlieb, and Guiso, 2017; Catherine, 2021). Other studies have documented heterogeneity in the returns to wealth and financial sophistication, showing that sophistication covaries with characteristics like education (Calvet, Campbell, and Sodini, 2007; Fagereng et al., 2020). We contribute to the literature by presenting evidence that the polygenic score we study can be related to financial sophistication, which can lead to compounding effects of genetic endowments over the life-cycle.

The remainder of this paper is organized as follows. Section 2 describes our data and presents motivating summary statistics. Section 3 presents our model and outlines our estimation strategy. Section 4 presents parameter estimates, the fit of the empirical patterns that we target, and the implications of the estimates. Section 5 uses the model to assess how socioeconomic outcomes, lifetime welfare, and their relationship with the EA score would change under two cost-saving changes to the Social Security system. Section 6 concludes.

2. Data

2.1. The Health and Retirement Study and Our Sample

Our empirical analysis uses data from the Health and Retirement Study (henceforth HRS). The HRS surveys a representative sample of more than 20,000 Americans over the age of 50 and their spouses. The longitudinal design of the survey features biennial waves starting in 1992 and continuing until the present, which provide information on respondents' labor supply, income, wealth, financial decisions, retirement, mortality, and inheritances. Retrospective survey questions ask about childhood socioeconomic status and parental characteristics.[3] The HRS data can also be linked to Social Security Administration records to provide data on earned income throughout the life-cycle. Crucially, the HRS contains genetic information on over 18,000 respondents collected from 2006 onward. Genetic data allow us to construct various measures of genetic endowments from the behavioral-genetics literature for HRS respondents, and to study their associations with other HRS variables. These measures include polygenic scores, which are summary indices of genetic variants that have been shown to predict observable outcomes.

This paper uses a polygenic score for educational attainment developed by Lee et al. (2018) as a measure of genetic endowments that influence various dimensions of human capital.[4] We refer to it as the *EA score*. This individual-level measure aggregates genetic variants that have been linked to educational attainment. It is standardized to have a mean of 0 and a variance of 1. A relatively high EA score

[3]The HRS also collects information on a host of factors that are omitted from our analysis, including variables on health and family structure.

[4]New polygenic scores are developed and updated for different traits as more data becomes available. In the case of educational attainment, Okbay et al. (2022) construct a polygenic score based on a larger discovery sample than that of Lee et al. (2018) (3 million vs. 1.1 million individuals) and which also improves upon its predictive power. We use the polygenic score of Lee et al. (2018) because it is the latest that has been calculated and made available for HRS respondents.

indicates that an individual possesses relatively more of the genetic variants that have been empirically linked out-of-sample (in our case in non-HRS samples) to educational attainment. Previous studies have shown that the EA score is predictive not only of completed education but also of other economic outcomes such as labor supply, income, and wealth, even after flexibly controlling for educational attainment (Papageorge and Thom, 2020; Barth, Papageorge, and Thom, 2020). Since earlier papers have discussed the EA score in depth, we refer the reader to Appendix A for more details. In that appendix we present an overview of how polygenic scores like the EA score are constructed, important limitations in their interpretation and use, and previous research in the social sciences that has used the EA score.

Data availability and issues surrounding the interpretation of polygenic scores detailed in Appendix A place several restrictions on our sample. We begin with the sample used in Barth, Papageorge, and Thom (2020), which includes only households with members of European ancestry. As explained in Martin et al. (2017), a set of technical issues means that the incorporation of non-European households into our analysis would be misguided and could generate misleading conclusions about cross-ethnic group genetic differences. The sample is also limited to households with non-missing data on key measures of interest for this study, including wealth, stock market participation, and Social Security Administration earnings records. These requirements permit a maximum sample size of 2,590 households (5,701 household-year observations) from the overall HRS sample of over 20,000 households and over 160,000 household-year observations.

We make further sample restrictions aligned to the structural model. In particular, the model is a unitary household model that abstracts from joint labor supply decisions and marriage dynamics. We focus on households that i) enter the HRS panel as married or partnered two-person male-female households; ii) remain intact (except due to the death of the female head of household); iii) are not observed earning income from jobs not covered by the Social Security Administration (SSA) data; iv) receive at least 70 percent of their SSA earnings income from the male partner; and v) have non-missing genetic data for the male partner, which we take to be the genetic endowment of the household. This set of restrictions generates a main analytic sample of 870 households with wealth data observed for a total of 2,318 household-year observations. We note, however, that to generate income and inheritance moments, we use slightly different samples. Samples used for the inheritance and income processes are described in Appendix B.1.

We use a comprehensive measure of household wealth that includes the net value of financial assets (cash, checking and saving accounts, certificates of deposits, stocks, bonds, mutual funds, trusts, and others, minus the value of non-housing debt); the net value of housing and businesses; and the balances of retirement accounts such as 401k and Keogh accounts. This is similar to the measure of wealth used in Barth, Papageorge, and Thom (2020) with the exception that we exclude the present value of Social Security payments, defined benefit pensions, and annuity income. We take the values of these components of wealth from the RAND HRS "Detailed Imputations" files. Our analysis also uses a binary indicator of whether households own any stocks, which takes a value of one for direct holdings, mutual fund holdings, and holdings through retirement accounts.

For our measure of labor income, we use the Respondent Cross-Year Summary Earnings data set of the HRS, which contains earnings data from the Social Security Administration's Master Earnings File (MEF). We use individuals' total earnings from the MEF, which include "regular wages and salaries, tips, self-employment income, and deferred compensation" (see Olsen and Hudson, 2009, for a detailed description of the components and history of measured earnings in the MEF). The earnings are derived from tax filings and are available in the Summary Earnings data set at an annual frequency, starting from the year 1951. Because the earnings data in the MEF were initially collected with the purpose of calculating Social Security benefits, earnings are top-coded at their maximum taxable level, which changes every year. As in Barth, Papageorge, and Thom (2020), we use data from the Current Population Survey to replace top-coded amounts with average earnings conditional on earning at least the maximum taxable amount for a given year.

Table 1 presents basic descriptive statistics for the analytic sample. Approximately 30 percent of the sample has a college degree, making the sample more highly educated than the overall HRS population. Retirement rates increase from approximately 17 percent for ages 50-62 to about 59 percent for ages 63-67. By age 73, 86 percent of the sample has retired. Table 1 also presents basic descriptive statistics on the log of total

Table 1: Summary Statistics.

A. Demographics, Income and Wealth				B. Childhood Socioeconomic Status (SES)			
Variable	Mean	SD	N	Variable	Mean	SD	N
Birth Year	1939.5	5.81	870	Mother's Educ.	10.52	2.97	766
College	0.30	0.46	870	Father's Educ.	10.20	3.58	766
Retired				*Family SES*			
50-62	0.17	0.37	3,350	Well Off	0.07	.	766
63-67	0.59	0.49	1,786	Average	0.67	.	766
68-72	0.78	0.42	1,393	Poor	0.24	.	766
73+	0.86	0.35	503	Varied	0.01	.	766
Total Prime-Age Income (× $1,000)				*Father's Job*			
Mean	1,984.48	.	870	Manager / Prof.	0.18	.	766
Std. Dev.	802.85	.	870	Sales	0.07	.	766
25th Percentile	1,451.90	.	870	Clerical	0.03	.	766
50th Percentile	1,928.22	.	870	Service	0.04	.	766
75th Percentile	2,456.86	.	870	Manual / Operators	0.64	.	766
Household Wealth (× $1,000)				Armed Forces	<0.01	.	766
Mean	716.34	.	2,318	Don't Know	0.03	.	766
Std. Dev.	926.29	.	2,318	Missing	<0.01	.	766
10th Percentile	52.12	.	2,318	*Child Health*			
25th Percentile	150.90	.	2,318	Excellent	0.55	.	766
50th Percentile	372.03	.	2,318	Very Good	0.27	.	766
75th Percentile	872.71	.	2,318	Good	0.13	.	766
90th Percentile	2,871.72	.	2,318	Fair	0.03	.	766
Any Stocks	0.68	0.47	2,318	Poor	0.01	.	766

Summary statistics for demographics, income, wealth, and SES variables in our main analytical sample.

prime-age SSA earnings. To construct this variable, we sum male earnings from age 30 to 60. The median of total income over this age range is $1.9 million, which would constitute an average of approximately $59,000 per year in 2010 dollars. Table 1 also provides detailed descriptive statistics for wealth for household-year observations in which the male household member was aged 60-70. The median wealth in the sample is approximately $372,000.

2.2. Genetic Endowments and Family Environments

A natural concern is that genetic endowments are endogenous to family environments. Parents who provide their children with genetic material also provide them with family environments, including resources that could benefit their educational attainment, labor market behaviors and outcomes, financial decision-making, and wealth accumulation. As many such factors are likely to be unobserved and thus omitted from empirical analyses, estimated coefficients relating the EA score to outcomes, including education, are likely to be upwardly biased. A host of studies have employed different methods to address this concern. For example, Trejo and Domingue (2018) and Belsky et al. (2018) rely on within-sibling variation in the EA score. This amounts to adjusting for a family fixed effect, and nearly all relationships hold despite some differences. Another method is to control for a rich set of variables that describe childhood environments. Both Ronda et al. (2022) and Arold, Hufe, and Stoeckli (2022) show that controlling for observed measures of family background can reduce the bias substantially. Since the HRS does not have the data to perform a within-family analysis, our approach is to incorporate information on childhood SES to control for key dimensions of childhood socioeconomic status. In what follows, we discuss how we summarize this information into a single variable.[5]

To construct a summary SES measure for the male members of each household in our sample, we estimate a cross-sectional regression of the following form:

$$\text{Educ}_i = b_0 + b_1 \text{EA}_i + b_2 X_i + e_i \tag{1}$$

Here X_i contains a vector of background variables that includes: mother's years of schooling, father's years of schooling, dummy variables for different subjective assessments of family SES growing up, dummy variables for categories of father's occupation growing up, and dummy variables for subjective assessments of health in childhood. Table 1 presents descriptive statistics on the components of X_i. After estimating the above equation, we construct an index of childhood SES as: $\text{SES Score}_i = \widehat{b}_2 X_i$. We normalize this measure so that it has a mean of zero and a unit standard deviation. While we do not have exogenous variation in the polygenic score EA_i in our sample, the results from Ronda et al. (2022) and Arold, Hufe, and Stoeckli (2022) give us some reason to believe that once we condition on SES Score_i, the associations we observe between EA_i and human capital outcomes may be close to causal effects.

2.3. Descriptive Associations

Table 2 presents basic regressions that highlight strong empirical relationships between the EA Score, wealth, and stock market participation, which help to motivate our structural model. Panel A presents regressions of log household wealth on

[5]In section 3, we specify the structural model and discuss how this variable enters through unobserved heterogeneity that is correlated with the EA score.

explanatory variables, including the EA Score for household-year observations where the male household member is aged 60-70.[6] Column (1) includes the EA Score, the SES Score, and a dummy variable for college education as controls. The coefficients on all three variables are substantial and statistically significant. The results suggest that a one standard deviation higher EA Score is associated with 27 percent higher wealth. Column (2) adds the log of total prime-age income, and Column (3) adds a dummy variable for holding any stocks in a given household-year. Controlling for income reduces the coefficient on the EA Score to 0.20, while additionally controlling for stocks reduces the coefficient to 0.15. However, this coefficient remains significant, suggesting that in this sample, even controlling for college, childhood SES, life-time earnings, and stocks, a one standard deviation higher EA score is associated with approximately 15 percent higher household wealth. Panel B of Table 2 presents regressions of a dummy variable for holding any stocks on the EA score, college, and childhood SES. The estimates in Column (4) suggest that a one standard deviation higher EA score is associated with a 6.8 percentage point increase in the likelihood of owning stocks. This coefficient is attenuated (0.049 v.s. 0.068) but remains highly statistically significant after controlling for lifetime earnings in Column (5).

The descriptive associations presented in Table 2 are consistent with earnings and portfolio choice playing major roles in mediating the relationship between the EA Score and household wealth. The model we develop and estimate below attempts to explain these associations as arising from the effects of the EA Score on the earnings process, the fixed costs of stock market participation, returns on stock market investments, and the disutilty of labor, while accounting for family background and the inheritance process. We focus on these channels in part because of empirical results found in Barth, Papageorge, and Thom (2020). However, several other channels could theoretically link the endowments measured by the EA Score and household wealth. In Appendix B, we explore five such mechanisms: business ownership, risk preferences, fertility, marital history, and longevity expectations. Measures of all of these channels significantly predict household wealth but only modestly attenuate

[6] To reduce omitted variables bias, it is common practice to control for the first 10 principal components of the full set of genetic variables, which helps to control for broad patterns in the genetic data that might arise from ethnic or regional differences and are omitted from the analysis. Appendix A discusses this practice in more detail.

Table 2: Summary Statistics: Mechanisms.

	Panel A: log Wealth			Panel B: Any Stocks	
	[1]	[2]	[3]	[4]	[5]
EA Score	0.276	0.203	0.152	0.068	0.049
	(0.054)	(0.050)	(0.045)	(0.015)	(0.014)
SES Score	0.294	0.222	0.169	0.069	0.050
	(0.054)	(0.058)	(0.052)	(0.016)	(0.015)
College	0.527	0.331	0.260	0.119	0.067)
	(0.109)	(0.105)	(0.095)	(0.032)	(0.030)
log Prime Inc.		0.969	0.699		0.255
		(0.195)	(0.180)		(0.030)
Any Stocks			1.058		
			(0.093)		
N	2259	2259	2259	2259	2259

This table reports results from models predicting log Wealth (Panel A) and a dummy variable for any stocks (Panel B). All regressions include the following controls: the first 10 principal components of the genetic data, an indicator for missing SES scores, and interactions between this missing SES indicator and any featured explanatory variables (the EA Score, College, log Prime Income, or Any Stocks if they are present in the specification).

the EA Score's associations with wealth and stock ownership. We thus abstract from these channels in developing our structural model.

3. Model

The model features heterogeneous agents that live from age 21 to a maximum age of 90. Each year, agents decide how much wealth to consume, how to allocate savings between a risky and a risk-free asset, and—once old enough—whether to retire or continue working for another year. Below we describe each component of the agents' dynamic problem.

3.1. Ex-Ante Heterogeneity

Agents are indexed by subscript i. They enter the model with four dimensions of observable heterogeneity: birth year (BY_i), an indicator for college completion (Coll_i), the EA score (EA_i), and an indicator equal to one if the agent participates in a defined benefit pension plan (DB_i). We allow these characteristics to influence agents' income, utility cost of work, expected returns on the risky asset (stocks), and the cost of stock market participation.

We also allow unobserved heterogeneity to affect agents' earnings, stock market participation costs, and expected stock returns. We model these three dimensions of heterogeneity as a vector of individual-specific "fixed effects" $\vec{\zeta}_i = [\zeta_i^w, \zeta_i^F, \zeta_i^R]'$. These parameters account for heterogeneity along dimensions that we do not model directly. One particularly important source of heterogeneity is childhood socioeconomic status and rearing environment. If the EA score in part represents genetic endowments that are conducive to achieving higher levels of income and financial sophistication, then we would expect the average-, low-, and high-EA individuals to grow up in different environments. This indirect channel, known as genetic nurture, could lead us to overestimate the influence of the EA score (Kong et al., 2018; Young et al., 2018; Ronda et al., 2022). To address this issue, we model unobserved heterogeneity as a combination of a fully random component ($\vec{\mathcal{Z}}_i$) and a component that is correlated with the EA score through childhood socioeconomic status (SES_i),

$$\underbrace{[\zeta_i^w, \zeta_i^F, \zeta_i^R]'}_{\vec{\zeta}_i} = \underbrace{[z_w, z_F, z_R]'}_{\vec{z}} \times \text{SES}_i + \underbrace{[\mathcal{Z}_i^w, \mathcal{Z}_i^F, \mathcal{Z}_i^R]'}_{\vec{\mathcal{Z}}}, \qquad \vec{\mathcal{Z}} \sim \mathcal{N}\left(\vec{0}, \Sigma_{\mathcal{Z}}\right) \tag{2}$$

where $\Sigma_{\mathcal{Z}} = \text{diag}[\sigma^2(\mathcal{Z}^w), \sigma^2(\mathcal{Z}^F), \sigma^2(\mathcal{Z}^R)]$, and $\text{SES}_i = \phi EA_i + \varepsilon_i$, with $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Here $\vec{z}$, $\Sigma_{\mathcal{Z}}$, ϕ and σ_ε^2 are parameters to be estimated.

3.2. Utility

A surviving agent i in period t derives utility from consumption and leisure through the utility function

$$u_{i,t}(C_t, \ell_t) = \frac{\left[C_t^\gamma \left(1 - 0.34 \times \ell_t\right)^{1-\gamma}\right]^{1-\omega}}{1-\omega} - d_{i,t} \times \ell_t, \tag{3}$$

where C_t is consumption and ℓ_t is a binary variable that indicates whether the agent is working ($\ell_t = 1$) or not ($\ell_t = 0$). This assumes that labor is indivisible and that it consumes 34% of an agent's endowment of time (8 of 24 hours per day), which we normalize to 1. A Cobb-Douglas function aggregates consumption and leisure with a weight $\gamma \in [0, 1]$ on consumption. The aggregate passes through a constant relative-risk-aversion function with coefficient of relative risk aversion ω. The term $d_{i,t} \times \ell_t$ is an additive utility cost of work. We allow this cost to vary with agents' education, EA score, and age:

$$d_{i,t} = d_0 + d_{\text{Coll}} \times \text{Coll}_i + d_{\text{EA}} \times \text{EA}_i + d_{\text{Age}} \times \max\{\text{Age}_{i,t} - 50, 0\}. \tag{4}$$

Our specification of the utility function allows for two different sources of heterogeneity in the disutility of work that could drive the empirical relationships between observable characteristics—the EA score in particular—and retirement patterns. The Cobb-Douglas aggregator in Equation 3 makes the value of leisure change with agents' consumption and, therefore, agents with different income levels (that support different consumption levels) will value leisure differently. This channel, whose importance is highlighted by Heckman (1974), is common in recent models of life-cycle labor supply (see, e.g., Low, Meghir, and Pistaferri, 2010; Blundell et al., 2016). The additive term in Equation 4 is intended to capture costs of work that are unrelated to the level consumption. The heterogeneity of this cost is intended to capture potential differences in the physical intensity of jobs that may co-vary with observable traits. Studies like Attanasio, Low, and Sánchez-Marcos (2008); Bagchi (2015) have used specifications with both additive and multiplicative costs of labor similar to ours.

Finally, to accommodate the fact that high-income households have higher saving rates and are slow to run down their wealth at old ages (Dynan, Skinner, and Zeldes, 2004), we include a "joy of giving" bequest motive (Carroll, 2002) that produces utility from end-of-life wealth. We use the functional form of De Nardi, French, and Jones (2010), in which a person who dies with savings $S_{i,t}$ receives utility

$$\varphi(S_{i,t}) = \theta \frac{(S_{i,t} + \kappa)^{1-\omega}}{1 - \omega}, \tag{5}$$

where ω is the same coefficient of relative-risk-aversion as above and θ and κ are parameters that we estimate and which govern the intensity of the motive and the degree to which bequests are a luxury good.

3.3. Labor Income

We model pre-tax labor income $\tilde{W}_{i,t}$ as the log-sum of a deterministic component that depends on individual characteristics and aggregate trends, the fixed unobserved heterogeneity draw $\mathcal{Z}_i^w$, and an agent- and time-specific shock $\epsilon_{i,t}^w$:

$$\ln \tilde{W}_{i,t} = f(\text{Age}_{i,t}, \text{EA}_i, \text{Coll}_i, \text{SES}_i, \text{DB}_i, \text{Year}_t, \text{Unemp}_t) + \mathcal{Z}_i^w + \epsilon_{i,t}^w, \tag{6}$$

where Unemp_t is the aggregate unemployment rate in period t. The wage shock is independent across time and agents and is normally distributed, $\epsilon_{i,t}^w \sim \mathcal{N}(0, \sigma^2(\epsilon^w))$. We present our specification and estimates for pre-tax labor income in Appendix B.1.

We use two types of labor income taxes: a constant-rate tax (as in Chai et al., 2011), and an additional proportional tax that applies up to a year-dependent maximum $\bar{T}(\text{Year})$ and represents Social Security taxes.[7] Given tax rates τ^W and τ^{FICA}, we model this tax scheme with a function $\tau_t(\cdot)$ that computes post-tax income as $\tau_t(\tilde{W}) = \tilde{W}(1 - \tau^W) - \tau^{\text{FICA}} \times \min\{\tilde{W}, \bar{T}(\text{Year}_t)\}$.

3.4. Retirement and Social Security

Once agents reach a minimum age, they can decide to retire. This decision is irreversible and happens after receiving labor income so that it takes effect in the following year. We impose the restriction that the first year of retirement has to occur in an age interval $[\text{Age}_0^R, \text{Age}_f^R]$, which we set to $[62, 80]$. Agents with defined benefit pension plans start receiving their payments in the first year in which they do not work. We assume that agents start claiming Social Security benefits the moment they stop working or at the minimum claiming age ($\text{Age}_{\min}^{SS}$) if they stop working before that.[8]

Consistent with the payout policy in the U.S., social security benefits are a concave function of agents' average income over their highest-earning years and increase with each additional year of work up to a maximum. Our methodology for computing benefits follows that of the Social Security Administration; we provide a detailed description in Appendix C. Our main simplification is that we use expected (rather than realized) earnings when computing SS benefits. That is, in each period t in which the agent could choose to retire, instead of the agent looking backwards at her *realized* earnings (the top 35 years of which determine SS benefits), the agent anticipates

[7] We use the Social Security Administration's historical maximum taxable incomes as $\bar{T}(\text{Year}_t)$.

[8] In our baseline scenario, the minimum claiming age and minimum retirement age are both 62, and therefore agents will always start claiming Social Security benefits the same year they stop working. This changes in our counterfactual policy experiments, one of which shifts the minimum claiming age but leaves open the possibility of agents retiring before that point.

receiving SS benefits that are determined by her *expected* top-35 yearly earnings as of period t, based on the income function described in Equation (6). This allows us to avoid including the additional state variable of realized cumulative earnings. Additionally, because the model has no permanent income shocks, transitory shocks over the life-cycle should largely cancel out, and expected earnings should be a good representative of realized earnings. We use $\text{SSB}_i(n)$ to denote the yearly benefits that person i would receive if she retired at age n.

In addition to social security benefits, retirees who have a defined benefit pension plan ($\text{DB}_i = 1$) receive pension income. We model the annual amount of DB pension payments (DBf_i) as a log-linear function of the time-invariant components of income, EA_i, and Coll_i, which we estimate from the data. This is similar to our treatment of social security income: to avoid the introduction of an additional state variable, we tie DB retirement benefits to the predictable component of earnings rather than realized earnings. We present the full specification and estimates of the defined benefit pension income process in Appendix B.1. Both social security benefits and defined benefit pension flows are taxed at a constant rate τ^s.

3.5. Inheritances

Every period, agents receive inheritances ($\text{Inher}_{i,t}$) that follow an agent- and age-specific stochastic process. With probability $P_i^I(\text{Age}_{i,t})$, individuals receive an inheritance in period t of amount $\text{C.Inher}_i(\text{Age}_{i,t})$. With probability $(1 - P_i^I(\text{Age}_{i,t}))$, the individual does not receive an inheritance in periods t. Both the probability of receiving an inheritance P_i^I and the value of inheritances conditional on reception C.Inher_i depend on Coll_i and EA_i. We estimate both functions using our HRS sample; their specification and parameter estimates can be found in Appendix B.1.

3.6. Financial Assets

At the end of each period, agents decide how to allocate their savings between two financial assets: a risk-free bond with return factor R and a risky asset representing the stock market with an agent- and time-specific return factor $\tilde{R}_{i,t}$. Short-selling of either asset is not permitted.

For agent i, the risky return factor follows the process:

$$\ln \tilde{R}_{i,t} = \ln R_t^{\text{SP500}} - \mu^{\text{SP500}} \times g(r_0 + r_{\text{Coll}} \times \text{Coll}_i + r_{\text{EA}} \times \text{EA}_i + \zeta_i^R), \tag{7}$$

where $\ln R_t^{\text{SP500}}$ and μ^{SP500} are the log-return of the S&P 500 stock market index in year t and its mean log-return, respectively. The factor $g(\cdot)$ captures agents' degree of inefficiency when investing in risky assets. It takes the form of a logistic function: $g(x) = \frac{e^x}{1+e^x}$, which ranges from 0 to 1. An efficient agent ($g \approx 0$) will replicate the market's returns. An inefficient agent ($g \approx 1$) will have expected log-returns close to 0. This parsimonious specification is similar to that of Lusardi, Michaud, and Mitchell (2017). Note that all agents face the same risk on the risky asset return; only the mean return is agent-specific.

We conceive of heterogeneity in risky asset returns as proxying for sound investment decisions. Paying higher fees on mutual fund investments or excessive trading (and resulting taxes), for instance, would degrade the average return earned in the market. Inopportune market-timing strategies, such as buying during periods of high price-earnings ratios and selling during periods of low price-earnings ratios, would also be detrimental to returns. Further, because risk is not agent-specific, lower expected returns also mean lower Sharpe ratios. From this perspective, a (roughly) equivalent interpretation of lower expected returns would be higher risk for a given level of expected return. Greater risk exposure conditional on expected returns would be consistent with poor diversification. Agents assume stock returns follow a normal distribution given by $\ln R_t^{\text{SP500}} \sim \mathcal{N}(\mu^{\text{SP500}}, \sigma^{\text{SP500}})$.

To own the risky asset, agents must pay a per-period monetary participation cost that represents the administrative and opportunity costs of managing investments (Vissing-Jorgensen, 2002). The cost F_i depends on *ex-ante* demographic characteristics:

$$\ln \mathrm{F}_i = f_0 + f_{\text{Coll}} \times \text{Coll}_i + f_{\text{EA}} \times \text{EA}_i + \zeta_i^F. \tag{8}$$

Capital gains are taxed at a constant rate τ^c.

3.7. Recursive Representation and Timing Summary

Agents maximize their expected discounted lifetime utility using a discount factor β and taking into account their probability of survival δ_t. An agent's possible choice variables at a given time are his consumption $C_{i,t}$, the fraction of his savings allocated to the risky asset $\phi_{i,t}$, and his retirement status for next period. His state vector consists of his beginning-of-period wealth $A_{i,t}$ and his current retirement status. Re-

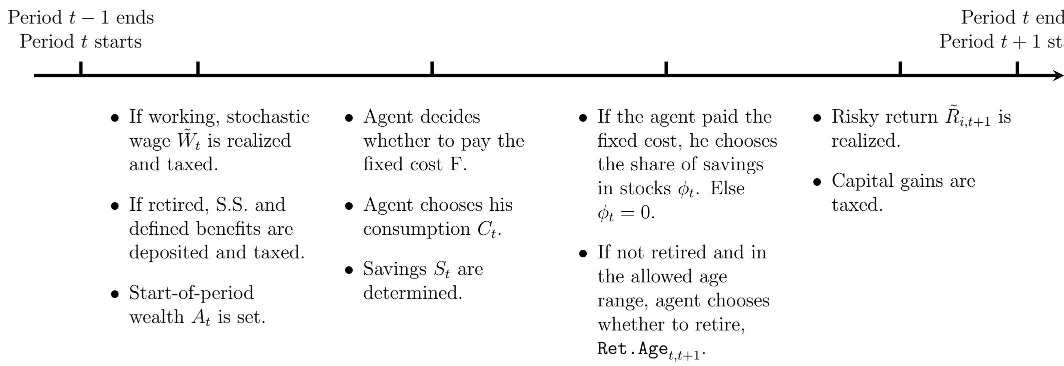


Figure 1: Timing of decisions and shock realizations.

tirement status is captured using the age at which the agent retired, $\texttt{Ret.Age}_{i,t}$. This discrete variable takes the value n if the agent retired at age n, and takes the null value $\emptyset$ if the individual has not yet retired. We track the age of retirement because it influences the level of social security benefits that agents receive. Figure 1 summarizes the timing of decisions and shocks in our model.

The choices and constraints that an agent faces depend on his age and retirement status. We illustrate the transitions of the model depicting the problem of an agent who has not retired yet ($\texttt{Ret.Age}_{i,t+1} = \emptyset$) but who has the option to retire. For a level of assets $A_{i,t}$, his value function is

$$
\begin{aligned}
V_{i,t}(A_{i,t}, \emptyset) = \max_{C_{i,t}, \phi_{i,t}, \texttt{Ret.Age}_{i,t+1}} \; & u(C_{i,t}, 1) + \beta \delta_t \mathbb{E}_t \left[V_{t+1}(A_{i,t+1}, \texttt{Ret.Age}_{i,t+1}) \right] + \not{\delta}_t \varphi(S_{i,t}) \\
0 \leq & C_{i,t}, \quad 0 \leq S_{i,t}, \quad 0 \leq \phi_{i,t} \leq 1, \quad \texttt{Ret.Age}_{i,t+1} \in \{\emptyset, \text{Age}_{i,t} + 1\}, \\
S_{i,t} = & A_{i,t} - C_{i,t} - \text{F} \times \mathbf{1}\left[\phi_{i,t} > 0\right], \\
A_{i,t+1} = & \left\{ (1 - \tau^c) \left[\phi_{i,t} \tilde{R}_{t+1} + (1 - \phi_{i,t}) R \right] + \tau^c \right\} \times S_t + \text{Income}_{i,t+1}(\texttt{Ret.Age}_{i,t+1}),
\end{aligned}
$$

where $S_{i,t}$ denotes the agent's savings and $\not{\delta}_t \equiv 1 - \delta_t$ is the probability of death. We have aggregated all sources of non-capital income for conciseness. We present disaggregated representations of all the optimization problems that agents solve at different points of their lives in Appendix D.

3.8. Estimation

Our approach to estimating model parameters relies on standard methods, so we relegate most details to Appendix B and provide a brief overview here. We estimate the model in two steps. In the first step, we directly estimate the specifications of wages (Equation 6), inheritances, and defined benefit pension flows on the HRS sample data. The details and results of the first step are found in Appendix B.1.

In the second step, we use the method of simulated moments (MSM) to estimate the parameters that govern the financial costs to stock market participation (Equation 8), inefficiencies in risky investments (Equation 7), the disutility of work (Equation 4), the dispersion of unobserved heterogeneity (Equation 2), the bequest motive (Equation 5), and the influence of childhood socioeconomic status on the unobserved heterogeneity draws for costs and returns (Equation 2).[9] The full set of parameters that we estimate internally is

$$\Theta = \{f_0, f_{\text{Coll}}, f_{\text{EA}}, \sigma(\mathcal{Z}^F), r_0, r_{\text{Coll}}, r_{\text{EA}}, \sigma(\mathcal{Z}^r), d_0, d_{\text{Coll}}, d_{\text{EA}}, d_{\text{Age}}, \theta, \kappa, z_F, z_R\}. \quad (9)$$

The algorithm proceeds as follows. A candidate set of parameters is chosen. Given these parameters, we solve the model, which delivers policy functions (mappings from state variables to choices) and transition rules (mappings from current-period state and choice variables to one-period-ahead state variables). Next, we simulate populations that match the HRS sample on observables. Given the observable dimensions along which agents can differ in our model, there are 400 potential types of *ex-ante* different agents.[10] However, only 190 of these possible combinations are actually observed in the HRS sample. For each person in our sample, we simulate 10 agents with matching characteristics. We simulate 27 such populations, one for each of the 27 potential combinations of draws of unobservable heterogeneity ($\vec{\zeta_i}$). Thus, our simulated populations have $5,130 = 190 \times 27$ types of *ex-ante* different agents. Armed with policy functions and transition rules, we can then simulate sequences of choices and outcomes, which delivers a data set that can then be compared to the actual

[9]Since income is observable, we estimate the parameters pertaining to unobserved heterogeneity of income (z_W, $\sigma(\mathcal{Z}^W)$) directly. See Appendix B.1.

[10]The observable dimensions of heterogeneity are birth year (BY_i), college completion (Coll_i), EA score (EA_i), and whether the agent participates in a defined benefit pension plan (DB_i). Appendix B discusses how we form the possible groups.

HRS analytic sample. To compare the simulated data set to the HRS data set, we compare some moments directly (e.g., the simulated and observed retirement wealth distribution). We also use indirect inference to identify parameters that are not directly observed. For example, we do not directly observe the disutility of labor. To identify this parameter, however, we can regress labor supply on observable variables in both the simulated and the HRS data set and then compare coefficients from these regressions. The moments that we directly target with our estimation routine are the mean and percentiles 10, 25, 50, 75, and 90 of the distribution of wealth between ages 60 and 70; the stock-market participation rate; and the coefficients of auxiliary regressions of log-wealth, stock ownership, and retirement on the EA score, education, our index of socioeconomic status, and other controls.

We repeat this process for many different candidate parameter sets and search for the parameter vector that minimizes the distance between the simulated and empirical moments. See Appendix B for a detailed description of the estimation procedure, non-estimated parameters, technical details, and a discussion of identification. Appendix E discusses the numerical solution of the model.

4. Results

4.1. Model Fit, Parameter Estimates, and Their Implications

Table 3 presents parameter estimates and their standard errors. To compute standard errors, we calculate the targeted moments for 50 different bootstrapped sub-samples of our analytical sample data. Within each subsample, we select the parameter vector from a pre-specified grid of 20,000 parameter vectors that minimizes the loss function between the selected vector the vector of targeted moments. This gives 50 sets of parameter vectors, one for each subsample, each of which minimizes the loss function within that particular subsample. We compute standard errors as the standard deviation of each parameter over the 50 resulting parameter vectors.[11]

Table 4 shows that the estimated model matches most of the empirical patterns remarkably well. We closely match the distribution of wealth up to the 75th percentile but under-predict wealth in the upper tail of the distribution. We also match relationships between the EA score, college attendance, wealth, and stock ownership as captured by the auxiliary regressions. The overall stock ownership rate is also close

[11]The 20,000 parameter vectors are the initial grid that we use in our main estimation routine.

Table 3: Internally-estimated parameters.

Participation cost
$\ln \mathrm{F}_i = f_0 + f_{\mathrm{Coll}} \times \mathrm{Coll}_i + f_{\mathrm{EA}} \times \mathrm{EA}_i + \zeta_i^F$

f_0	f_{Coll}	f_{EA}
-0.9867	0.0311	0.0066
(0.2092)	(0.0369)	(0.0143)

Risky asset returns
$\mathrm{Inefficiency}_i = g(r_0 + r_{\mathrm{Coll}} \times \mathrm{Coll}_i + r_{\mathrm{EA}} \times \mathrm{EA}_i + \zeta_i^R)$

r_0	r_{Coll}	r_{EA}
-0.0366	-1.1055	-0.6610
(0.0608)	(0.2568)	(0.1326)

Disutility from work
$d_{i,t} = d_0 + d_{\mathrm{Coll}} \times \mathrm{Coll}_i + d_{\mathrm{EA}} \times \mathrm{EA}_i + d_{\mathrm{Age}} \times \max\{\mathrm{Age}_{i,t} - 50, 0\}$

d_0	d_{Coll}	d_{EA}	d_{Age}
0.3961	-0.0052	-0.0033	-0.0241
(0.0816)	(0.0015)	(0.0009)	(0.0062)

Unobserved heterogeneity
$\vec{\zeta}_i = \vec{z} \times \mathrm{SES}_i + \vec{\mathcal{Z}}_i$

$\ln \sigma(\mathcal{Z}^F)$	$\ln \sigma(\mathcal{Z}^r)$	z_F	z_R
1.1838	-4.0326	-0.0434	-0.7250
(0.5698)	(1.3845)	(0.0475)	(0.1451)

Bequest motive
$\varphi(S_{i,t}) = \theta(S_{i,t} + \kappa)^{1-\omega}/(1-\omega)$

$\ln \kappa$	$\ln \theta$
7.0638	6.9423
(0.2474)	(0.5353)

This table presents parameter estimates from the method of simulated moments. See the main text for details about the model and targeted moments. Standard errors are reported in parentheses and calculated using a bootstrap approximation that is also discussed in the main text.

Table 4: Targeted moments in the HRS and in the estimated model.

Wealth distribution
Thousands of dollars

	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}	Mean
Data	52	151	372	873	1819	716
Model	52	166	427	822	1379	605

Wealth regression
Coefficients from: $\ln \mathrm{Wealth}_{i,t} = X_{i,t}\beta + \varepsilon_{i,t}$

	EA	SES	Coll	Stocks
Data	0.15	0.17	0.26	1.06
Model	0.17	0.18	0.26	1.12

Stock-ownership regression
Coefficients from: $\mathrm{Stocks}_{i,t} = X_{i,t}\beta + \varepsilon_{i,t}$

	EA	SES	Coll	ln E. Inc.
Data	0.049	0.050	0.067	0.255
Model	0.046	0.052	0.069	0.220

Retirement regression
Coefficients from: $\mathrm{Retired}_{i,t} = X_{i,t}\beta + \varepsilon_{i,t}$

	EA	Coll
Data	-0.0059	-0.0564
Model	-0.0059	-0.0575

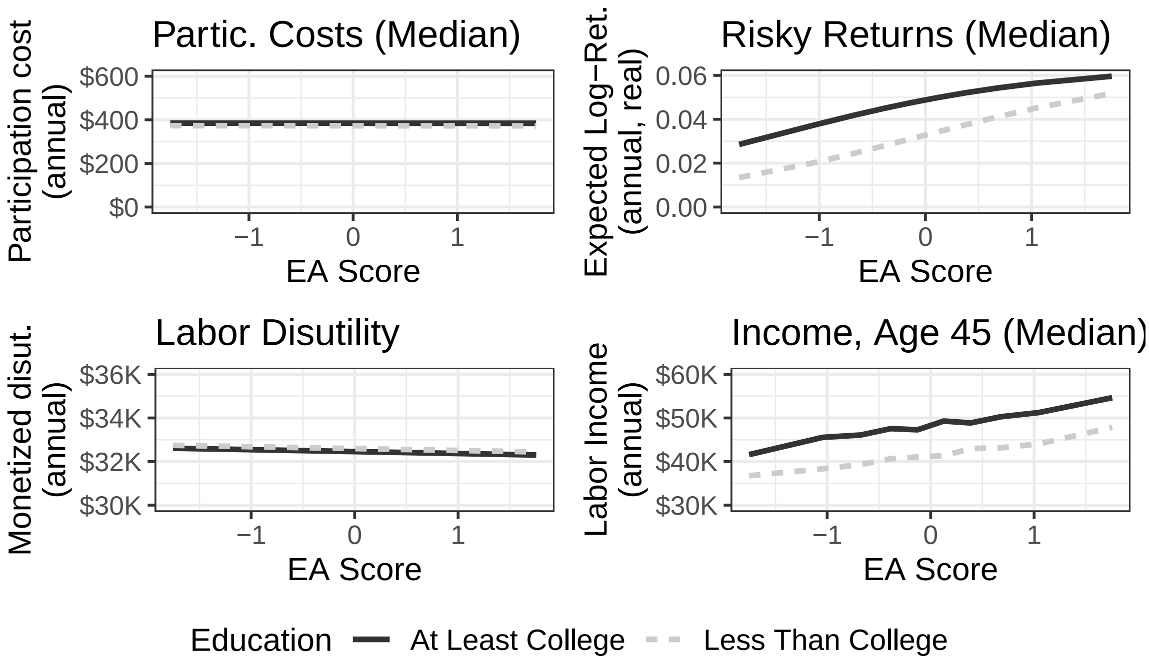
Retirement rates
Fraction of individuals retired by age bracket

	$[50, 62]$	$[63, 67]$	$[68, 72]$	≥ 73
Data	0.17	0.59	0.78	0.86
Model	0.11	0.69	0.91	1.00

Stock ownership rate
Ages 60 to 70

	Rate
Data	0.68
Model	0.72

This table reports the moments that we target in estimation, calculated both using our HRS analytical sample and using simulations from the model with the estimated parameter values. See the main text for variable definitions, sample descriptions, and complete specifications of the controls included in $X_{i,t}$ for each regression.



The disutility of labor is taken at age 50 and monetized (see the main text for details). We find the participation costs, expected log returns, and disutility from work at age 50 for each of our 5,130 agent types and weight them by the number of agents of each type in our simulated population. The figure presents medians at different levels of the EA score and educational attainment. Labor disutility depends only on agent's age, EA score, and educational attainment, so there is no remaining heterogeneity after conditioning on these characteristics. For income, we report the median of EA score-education combinations at age 45 from our simulations.

Figure 2: Economic fundamentals, the EA score, and education.

to the sample rate. For retirement, while the model fits the auxiliary regression that relates retirement to the EA score and college attendance, the age-binned retirement rates show slight discrepancies arising mainly from over-estimating the fraction of agents that retire between ages 63 and 67.

According to estimates in Table 3, both the EA score and education have negligible effects on the cost of stock market participation, F_i. This is confirmed by Figure 2, which plots the median participation costs, expected risky returns, disutility of labor, and income at age 45 against the EA score for both levels of education. The top-left panel shows no meaningful variation of the median annual participation cost with either the EA score or education. We estimate this cost to range between \$372 and \$385, which is similar to estimates in recent studies in household finance (see e.g., Fagereng, Gottlieb, and Guiso, 2017; Catherine, 2021).

Risky asset returns, however, appear to be significantly affected by both the EA score and education. Table 3 and the top-right panel of Figure 2 show that, between the lowest and highest EA deciles, the median expected log-return on stocks increases from 1.3% to 5.2% for those without a college degree and from 2.8% to 6% for those with a college degree. The direction of these estimated relationships is consistent with past studies (e.g., Calvet, Campbell, and Sodini, 2007; Fagereng et al., 2020), which find that returns to wealth are heterogeneous and correlated with wealth, income, and education, each of which covaries with the EA score.[12] However, part of the relationship between the EA score and returns that is depicted in Figure 2 is due to the relationship between the EA score and SES.[13] To evaluate the size of the estimated association that is not due to SES—the part that operates through r_{EA} in Equation 7—we calculate expected risky log-returns of our simulated population of agents assuming their EA scores were one standard deviation (1.0) higher, holding their other characteristics, including SES, constant. We find that, on average, the expected risky log returns of our agents would increase by 78 basis points (0.0078).

The estimated degree of heterogeneity in returns is plausible and broadly consistent with the findings of studies that have measured the distribution of risk-compensated returns in other countries or modeled it as an endogenous investment. By design, the functional form of the expected log-return of the risky asset (Equation 7) forces it to be between 0 and the benchmark μ^{SP500} for every agent. A similar range is used in studies such as Lusardi, Michaud, and Mitchell (2017).[14] Our estimates determine the distribution of simulated agents across this predetermined range. Appendix G presents detailed statistics of the distribution of risk-compensated returns implied by our estimates. Households in our estimated model receive risk compensations (measured by the Sharpe ratio) that can range from 36% of that of the market benchmark in the 25th percentile to 76% in the 75th percentile. These ranges are plausible when compared with those measured by, e.g., Calvet, Campbell, and Sodini

[12] As discussed in Section 3, our specification of risky returns implies that agents receive different compensation for taking financial risk. One way to see this is that they have different Sharpe ratios. We present the distribution of available Sharpe ratios that our estimates imply in Appendix G.

[13] Individuals with higher EA scores have a higher expected SES (see Section 3.1). This shifts their expected draws of unobserved heterogeneity (see Equation 2) and, in turn, their expected returns.

[14] A minor difference is that the lower limit for expected log-returns in Lusardi, Michaud, and Mitchell (2017) is set to the log risk-free rate instead of 0.

(2007) in Sweden and Gaudecker (2015) in the Netherlands. They are also consistent with the implications of the model developed by Lusardi, Michaud, and Mitchell (2017) in which, at age 50, the average agent earns between 43% and 67% of the market's log-return premium, depending on his education.

Similar to previous studies, our income estimates imply a positive and significant association between the EA score and labor income even after controlling for educational attainment. The bottom-right panel of Figure 2 depicts the median labor income of agents with different EA scores and levels of education at age 45, which is close to their peak for those without a college degree. At this age, the median earnings for those without a college degree in the 4th and 7th decile of the EA score are \$40,700 and \$43,000, respectively. For those with a college degree the median earnings at this age and the same EA score deciles \$47,500 and \$48,900 respectively. Our estimates of the income process are reported in Appendix B.1.

The estimated additive component of disutility from labor varies little with education, the EA score, and age. We monetize the disutility to convey its magnitude. For a utility cost $d_{i,t} > 0$, we find the monetary value $m_{i,t}$ that an agent with no dynamic considerations and a baseline consumption of \$40,000 would be willing to pay to avoid the cost $d_{i,t}$.[15] Formally, $m_{i,t}$ solves

$$u_{i,t}(C_t = \$40k, \ell_t = 1) = u_{i,t}(C_t = \$40k - \$m_{i,t}, \ell_t = 1) + d_{i,t}.$$

The bottom left panel of Figure 2 displays monetized costs of work $m_{i,t}$ at age 50 against the EA score for both education groups. The figure confirms that there is no meaningful heterogeneity in our estimated disutility of work.

The disutility of work parameter would most clearly affect retirement. If wealth and expected retirement income are held constant, a greater disutility of work implies earlier retirement rates. The null results for the disutility of work suggest the model can fit the observed retirement patterns based on variation in wealth, income, and expected retirement income, without needing to assign additional power to a distaste for work. This finding is also supported by the strong fit of the retirement regression moments.

[15] Our specification of the utility cost (Equation 4) is age-dependent. We base our calculations on its values at age 50.

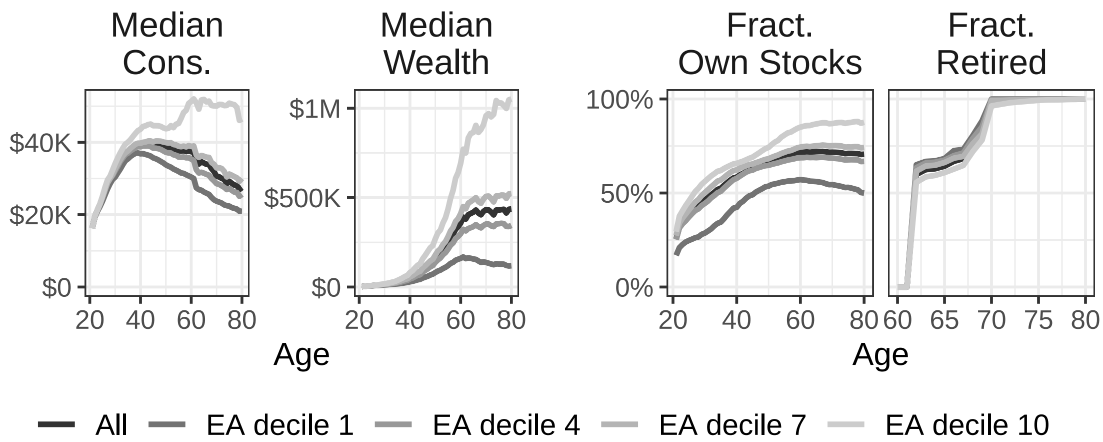


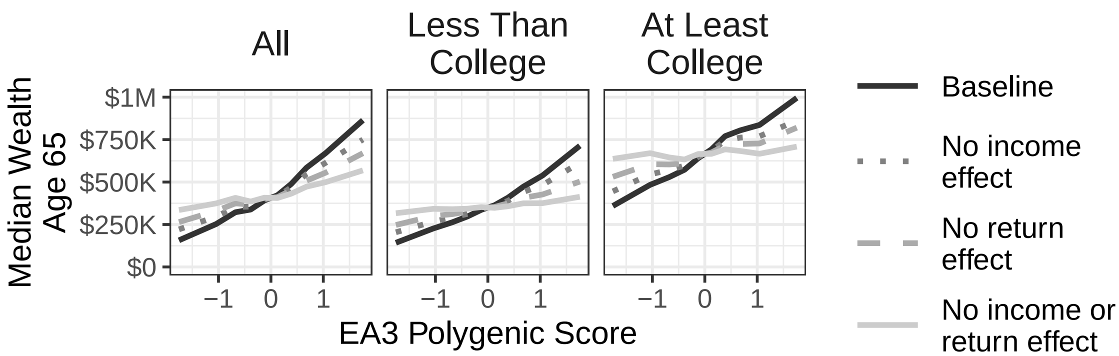
Figure 3: The choices and economic outcomes of agents with different EA scores.

4.2. Life-Cycle Choices and Outcomes

The estimates suggest no meaningful relationship between the EA polygenic score and either stock market participation costs or disutility from work but do suggest sizable associations with labor incomes and risky-asset returns, even at comparable levels of education. This section examines how these differences influence other choices and outcomes such as consumption, retirement, stock-market participation, and wealth across the life cycle. Figure 3 presents the simulated age-profiles of these choices and outcomes for agents in different deciles of the EA polygenic score.

The greater average incomes and more efficient investments of agents with higher EA scores afford them higher consumption throughout their lives. Differences in median consumption become noticeable around age 35. By age 40, the difference between the median consumption of agents in the 4th and 7th deciles of the EA score reaches $970 per year; this gap grows to $3,495 by age 60 and to $4,427 by age 80.

Despite higher consumption, agents with higher EA scores accumulate more wealth. Figure 3 shows that an EA score-wealth relationship emerges early in life and widens as agents age. The median wealth balances of agents in the 4th and 7th decile of the EA polygenic score are $45,175 and $56,928 at age 40, $297,997 and $410,610 at age 60, and $344,424 and $523,125 at age 80, respectively. The growth in wealth differences is due to the combination of higher labor earnings and higher returns on



"Baseline" corresponds to conditional medians calculated using the simulations of our estimated model. "No income effect" corresponds to simulations from a version of the model in which we set the EA score's coefficients in our income specification to 0, leaving the other components of the model unchanged. "No return effects" sets the EA score's coefficient on expected risky-asset log-returns to 0. "No income or return effect" sets both sets of coefficients to 0.

Figure 4: The EA score-wealth gradient and its sources.

savings, which compound to magnify differences at later ages. As Figure 3 shows, stock market participation is monotonically increasing in the EA score and peaks at age 67 for those with EA scores in the 4th and 7th deciles, reaching 69.0% and 75.5%, respectively.

Despite the lower levels of wealth and similar levels of labor disutility, differences in the retirement decisions of agents with different EA scores are small. Agents with higher EA scores retire at later ages on average. Our model matches this fact: 63.5% of agents in the 4th decile of the EA score retire at the minimum age of 62, and 71.3% have retired by age 67; for the 7th decile of the EA score, these numbers are 61.6% and 69.5%, respectively. There are a few possibilities that could explain this finding. First, lower EA score individuals will receive retirement income that represents a higher fraction of their lifetime earnings due to the progressivity of the social security system. Second, the labor earnings lost in retirement are lower for low EA score agents because they earn less income in the labor market on average. These two effects seemingly dominate the lower levels of wealth and consumption in retirement also experienced by low EA score agents.

The foregoing discussion highlights how the estimated effects of the EA score on labor income and financial proficiency compound to generate differences in wealth, stock market participation, and retirement. Among these outcomes, wealth has received the greatest attention, as a vast literature in economics has worked to understand key sources of wealth disparities. Barth, Papageorge, and Thom (2020) demonstrate that there is a robust association between the EA score and wealth at retirement. The model developed here allows us to disentangle multiple potential channels of wealth accumulation and to evaluate their importance in driving this relationship. A first driver of the relationship between the EA score and wealth is education. The EA score was built with the purpose of predicting educational attainment and therefore agents with higher EA scores will, on average, have more years of education that translate to greater earnings and wealth. After controlling for education by grouping individuals with a college degree and no college degree, the difference in median wealth at age 65 between the 1st and 10th EA score deciles falls from an unconditional value of \$707,423 to \$571,332 for those without a college degree and \$636,537 for those with a college degree.

In Figure 4, we further decompose the effects on wealth while continuing to control for education. To do so, we examine how the relationship between the EA score and wealth would change under different environments. First, we solve and simulate our model under an environment in which the income specification has all coefficients that multiply the EA score to 0, leaving the rest unchanged. The resulting wealth gradients are labeled "No income effect." The differences in median wealth between agents in the highest and lowest EA score deciles fall to \$409,547 for those without a college degree and \$416,348 for college graduates. Restoring the income process and eliminating the direct effect of the EA score on returns by setting $r_{\text{EA}} = 0$ in Equation 7 results in the wealth gradient labeled "No return effect." The differences in median wealth now fall to \$255,572 for those without a college degree and \$289,576 for college graduates, a significantly greater reduction than the results from removing the direct income effect. Finally, we remove the direct effects of the EA score on both income and returns. This reduces the gaps in median wealth between the top and bottom EA deciles to \$96,612 for those without a college degree and \$72,090 for college graduates, just 16.9% and 11.3% of the baseline difference. The small remaining difference is due to the combination of other direct effects, such as the EA score effect on the disutility of work and participation costs, as well as indirect effects, including the correlation of the EA score and childhood socioeconomic status, the influence of the

EA score on the inheritance process, and covariation between the EA score and other characteristics such as birth year and defined benefit pension arrangements.

5. Policy Experiments

The previous section demonstrated that genetic endowments are associated with multiple dimensions of wealth accumulation, including risky asset returns and labor market earnings. Crucially, the extent to which these differences affect individual outcomes and welfare may depend on the policy environments individuals face. Alternate policy environments may attenuate or magnify the association between genes and wealth over the life-cycle. Policies that strengthen or weaken the consequences of specific advantages (e.g., higher investment efficiency) will affect wealth inequality and genetic gradients. In this Section, we use our estimated structural model to examine the consequences of two counterfactual reforms to the Social Security system and study how the different policy environments affect behavior, outcomes, and welfare, in part through their interaction with genes.

While the existing literature on $G \times E$ interactions has sought to document the extent to which gene-outcome associations differ across different environments, these studies exclusively rely on environments that *have existed* and for which data are available. Instead, our estimated model permits analyses of gene-by-outcome associations in *counterfactual* environments that may not exist and possibly never will. This is critical for assessing the effects on genetic gradients across various policies options, of which only a small set may be chosen.

Specifically, we analyze two counterfactual changes to the Social Security system, which is one of the most consequential programs for elderly adults in the United States. Social Security is a federal program that provides monthly income to retirees based on retirement age and lifetime earnings. In response to the fiscal challenges to Social Security spurred by an aging population, the past few decades have seen many proposals to reform the system. Indeed, multiple proposals have been presented every year.[16] To reduce the cost of the Social Security system, a policy would have to collect more resources from young workers, reduce the benefits distributed to retirees, or involve some combination of the two. Such changes can affect welfare on multiple

[16] The Office of the Chief Actuary of the Social Security Administration maintains a repository of fiscal analyses of the proposals presented by Congress (Social Security Administration, 2022a).

dimensions, but, as the primary support tool for low-wealth retirees, changes to the system may disproportionately harm already vulnerable populations. We study how potential policy changes could modify $G \times E$ interactions; specifically, the extent to which changes in the social safety net affect the link between the EA score, economic outcomes, and welfare.

We consider two cost-saving policies, each consistent with common proposals for preserving the Social Security system. The first policy, which we call the "benefit shift," is a five-year forward shift in the Social Security benefit schedule. The minimum benefit-claiming age increases from 62 to 67, and the benefit calculation formulas move forward by five years. If agent i retiring at age a currently receives yearly benefits equal to $\mathrm{SSB}_i(a)$, under the benefit shift policy he receives $\mathrm{SSB}_i(a-5)$. This policy mimics many proposals suggesting increases in the age of benefit eligibility to encourage people to work longer. Net costs decrease because people will either work for longer (and reduce the number of years in which they claim benefits) or retire before they become eligible to claim full benefits and receive lower benefits than they would in the baseline scenario. Importantly, we assume agents are aware of this policy starting at age 21 (when the model begins), and we do not model uncertainty about policy regimes.

The second policy, which we call the "benefit reduction," represents a broad class of provisions that would alter the way benefits are calculated, lowering the annual payments to some or all future retirees. We analyze a simple policy of reducing all benefits by a constant fraction ψ. If agent i retiring at age a receives benefits $\mathrm{SSB}_i(a)$ in the baseline scenario, this policy would make their benefits $(1-\psi)\mathrm{SSB}_i(a)$. To make the two policy changes comparable in fiscal terms, we calculate the reduction rate ψ that would produce the same increase in revenue as the previous policy change, accounting for endogenous behavioral responses. Details of this calculation are found in Appendix F. We find that the reduction that makes the two policies fiscally equivalent is $\psi = 0.2789$. As with the benefit shift policy, agents are fully aware of the benefit reduction policy when they enter the model.

While the two alternative policies are calibrated to generate the same aggregate revenue, they differ in the channels along which they incentivize individuals to adjust, and this generates different behavioral responses. In a model with complete markets

We simulate identical populations of agents that experience identical sequences of shocks under the status quo and the two alternative social security changes. We find the relative consumption and wealth changes with respect to the baseline scenario agent-by-agent and report the medians of those changes at every age. We also find the fraction of agents who are retired and who are holding stocks at every age and report the differences with respect to the baseline scenario.

Figure 5: Life-cycle adjustments to Social Security changes.

and without borrowing constraints, the response of an agent to two different social security policy environments that have the same present discounted value of benefits associated with each age of retirement would be the same. However, our model features incomplete markets, uninsurable income shocks, and borrowing constraints. Furthermore, the benefit-reduction and benefit-shift policies do not have the same present discounted value of benefits associated with each age of retirement: the shift has stronger monetary incentives for agents to extend their working lives. These features combine to generate different optimal changes in behavior in response to each policy. The different responses are evident in Figure 5, which depicts the effects of both policies on retirement, stock-ownership, consumption, and wealth, by age.

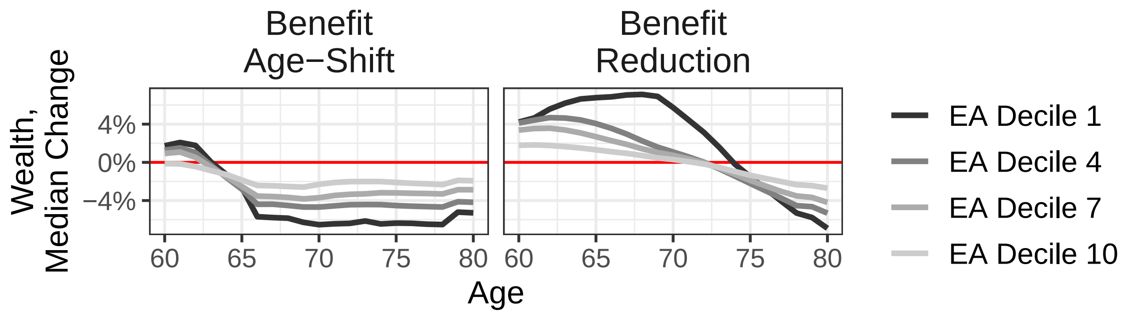
In response to the benefit-shift policy, agents delay their retirement and reduce their consumption. Figure 5 shows that, for example, the share of agents retired at age 70 is more than 30 percentage points lower than in the baseline policy environment. Even though some agents work longer, which implies higher labor income at later ages, the policy change leads to a small net decline in consumption that reaches roughly one percentage point at age 62 for the median agent. The reduced consumption generates a slight increase in wealth during the years before retirement, followed by a median decline of nearly four percentage points starting at age 62, which is driven by agents who retire but do not yet receive benefits. Stock-ownership patterns remain largely unchanged.

The benefit-reduction policy generates a smaller delay in retirement and a greater reduction in consumption than the benefit-shift. The difference in retirement rates with respect to the baseline environment peaks at age 68 with only an 11 percentage-point reduction. There is instead a greater and earlier reduction in consumption which increases in magnitude as agents age; by age 70, the median agent reduces his consumption by 2.5 percentage points. This larger shift in consumption means that, under the benefit reduction, wealth increases more dramatically than under the benefit-shift policy. This increase begins at age 30 and peaks at nearly 4% at age 62. After that, the benefit reduction leads to a rapid decline in wealth so that, by age 80, the median agent has 4% less wealth than under the baseline.

Our estimated model features rich heterogeneity in the costs and advantages that different agents have for changing their behavior in response to the two policies. This heterogeneity can produce differences in the ways that they chose to adapt and in the welfare losses that they ultimately bear. We examine these differences in the following subsection.

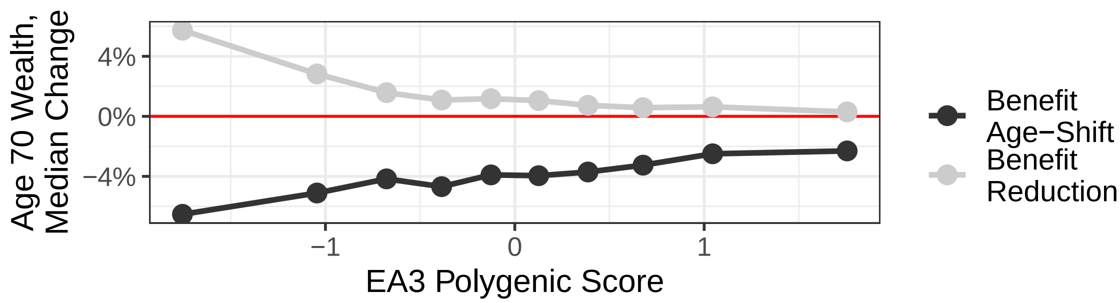
5.1. Gene-Outcome vs. Gene-Welfare Gradients

This section uses our estimated model to analyze how different social-security policy reforms would alter the association between the EA score and retirement wealth. The recent literature on $G \times E$ interactions has examined how changes in environmental factors alter the association between genetic endowments and outcomes of interest. However, rather than being restricted to realized environmental differences, this exercise highlights how structural models can be used for *ex-ante* evaluations of $G \times E$ interactions under *counterfactual policy environments*. The model also delivers estimates of the welfare impacts of the reforms for individuals with different genetic endowments. This is an important feature because, as we show, the gene gradient of a desirable outcome (e.g., retirement wealth) and the gene gradient of welfare can move in opposite directions.



We simulate identical populations of agents that experience identical sequences of shocks under the status quo and the two alternative Social Security changes. We find the relative wealth changes with respect to the baseline scenario agent-by-agent and report the medians of those changes at every age for different levels of the EA score.

Figure 6: Wealth response to policy changes by EA polygenic score.



We simulate identical populations of agents that experience identical sequences of shocks under the status quo and the two alternative Social Security changes. We find the relative wealth changes at age 70 with respect to the baseline scenario agent-by-agent and report the medians of those changes for different levels of the EA score.

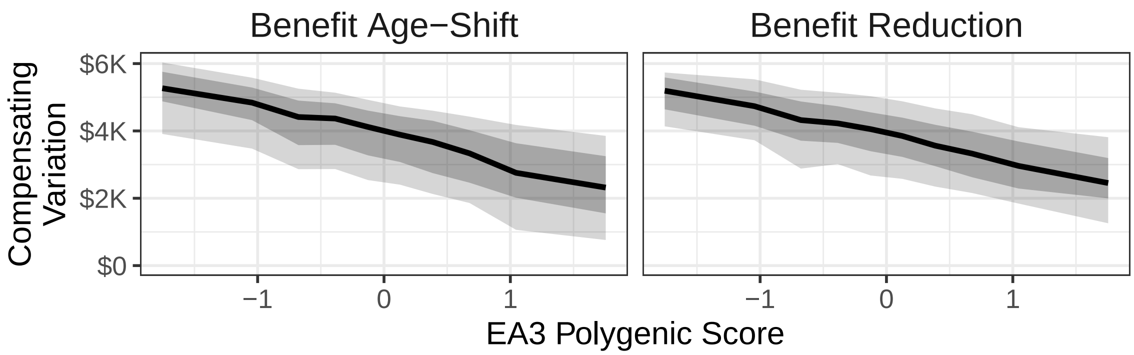
Figure 7: Social Security changes and the EA-wealth gradient.

The responses of lower EA score individuals to the policy changes produce greater relative adjustments in their retirement wealth than those of their higher EA score counterparts. Figure 6 shows the percentage changes in wealth that the two policies generate for individuals in different deciles of the EA score distribution by age. At age 70, the reductions in retirement wealth caused by the benefit-shift policy are more than twice as large for individuals in the lowest EA score decile (median of -6.5%) than for those in the highest EA score decile (median of -2.3%). By this same age, the increases in retirement wealth caused by the benefit-reduction policy are much larger for those in the lowest EA score decile (median of $+5.7\%$) than for those in the highest decile (median below $+1\%$).

The policy changes affect the empirical relationship between the EA score and retirement wealth through the differential responses of agents with different EA scores. Figure 7 presents the relative changes in wealth induced by both policies across the EA score distribution at age 70. The figure demonstrates that the benefit-shift policy steepens the EA score-wealth gradient because percentage reductions in wealth are greater for low EA score individuals and that the benefit-reduction policy flattens the gradient because increases are greater for low EA score individuals. The counterfactual exercise demonstrates changes in policies with identical fiscal consequences can flatten or steepen the relationship between the EA score and wealth. Gene-outcome gradients are not immutable "laws of nature" but are functions of policy environments and endogenous behaviors, and a well-posed economic model can deliver predictions about how gene-by-outcome associations may change under different policy regimes. And, perhaps most importantly, the model gives economic content to the sources of changing gene-by-outcome interactions.

If a person's genes are viewed a form of luck (Harden, 2021), then the adoption criteria of various policies may include whether genetic gradients are steepened or flattened. Under this criterion—and knowing that both policies raise the same aggregate revenue—the benefit-reduction policy may be naively considered more desirable than the benefit-shift, since the former flattens the genetic gradient in retirement wealth while the latter steepens it. However, ranking the two policies on the basis of their implications for the gene-wealth gradient is perilous. All else equal, wealth at a fixed age (say 70) may be a suitable proxy for overall welfare. However, *changes* in wealth at age 70 under different policy regimes may be a poor proxy for changes in welfare because the underlying economic conditions have also changed.

To analyze the welfare effects of the alternative policies for different individuals, we calculate the monetary amounts by which agents would need to be compensated to preserve their welfare after the policy reforms. We calculate these compensating variations for agents who are informed of the policy changes at age 21. If $V_{i,21}(\cdot)$ is the baseline value function of agent i at age 21 and $V^x_{i,21}(\cdot)$ is his value function under the alternative policy scenario x, the compensating variation from an initial wealth of W, $\mathrm{CV}^x_i(W)$, solves $V_{i,21}(W) = V^x_{i,21}\left(W + \mathrm{CV}^x_i(W)\right)$. That is, $\mathrm{CV}^x_i(W)$ is the monetary transfer that would be required to restore person i's expected lifetime welfare to what it was before the policy change; it will be higher for policies that bring greater reductions to i's lifetime welfare. These relationships implicitly define the compensating variation for a given policy as a function of the agent's characteristics—expected income path, preferences, financial proficiency, etc.—and his wealth W. We calculate the compensating variations for both policies for agents that match the characteristics of the sample that we use for estimating the model. We set the initial wealth of every agent to $W = \$20,000$.



The compensating variations for both policies are computed at age 21 and from a starting wealth of \$20,000. We find the compensating variation for each of our $5,130$ agent types and weight them by the number of agents of each type in our simulated population. The figure depicts percentiles of the distribution of these compensating variations at different levels of the EA score. The solid line corresponds to the median. Inner shaded areas cover observations between the 25th and 75th percentiles. Outer shaded areas cover observations between the 10th and 90th percentile.

Figure 8: Distribution of compensating variations at different levels of the EA score.

In spite of their different implications for the EA score-wealth gradient, both policy changes *steepen* the EA score-welfare gradient, as they generate greater welfare shortfalls for individuals with lower EA scores. Figure 8 plots the distribution of compensating variations by EA score for each policy. Despite differences in the margins along which agents modify their behavior in response to each policy, the figure reveals that the two policies have similar effects on the *welfare* of agents. The unconditional distribution of the compensating variations of both policies is similar: the mean and standard deviation for the benefit shift policy are \$3,780 and \$1,300 respectively, and for the benefit reduction policy, they are \$3,820 and \$1,100. However, there are large differences in the welfare costs borne by agents with different EA scores. For both policies, the median compensating variation for agents in the lowest EA score decile is more than twice as large as that for agents in the highest EA score decile. By generating greater welfare losses for agents with lower EA scores, both policies steepen the EA score-welfare gradient. Therefore, comparing the effects of different policies based on the relationship between genes and an outcome like wealth may obscure their effects on inequality of welfare.

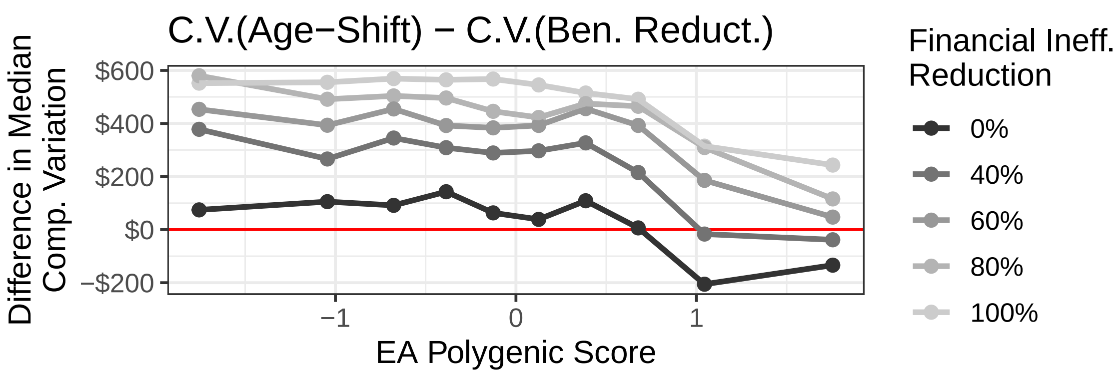
This exercise demonstrates that a model of economic behavior can be helpful for capturing the multiple factors that influence gene-outcome gradients and for analyzing the welfare implications behind them. Using our estimated model, we have simulated the effects of counterfactual policy reforms, taking into account behaviors that determine outcomes of interest, such as wealth and retirement, which change in response to the policies. We have also used the model to demonstrate that the effect of an environmental change on the relationship between genetic endowments and a desirable outcome may poorly reflect its impact on the relationship between genetic endowments and welfare. Failing to account for these behavioral responses and the complex relationship between observable outcomes and welfare can lead to flawed conclusions about the distributional consequences of different policies.

5.2. Investment Efficiency and Welfare Changes

The foregoing analysis demonstrates that in spite of the apparent differences in the distributional impacts of the two policies, both have similar welfare implications and, in particular, both steepen the EA score-welfare gradient. However, the behavioral responses and mechanisms highlighted by our model suggest further interventions that could moderate the impact of the policies on gene-welfare gradients. In this section, we analyze how interventions that increase the financial proficiency of our simulated agents can mitigate the impact of the Social Security reforms.

Our estimated model suggests that there are large differences in the barriers that agents face for accessing sophisticated financial products and in how effectively they use these products once they gain access. In an environment where the Social Security system becomes less generous and households become more reliant on their saving and investment decisions to support their retirement years, these differences become more important. A growing literature has evaluated the effectiveness of different types of interventions aimed at improving the financial knowledge and financial behavior of different groups of the population, which could potentially reduce these differences (Kaiser and Menkhoff, 2017; Kaiser et al., 2022).

We consider a scenario in which a successful intervention has dramatically improved the financial proficiency of our agents, lowering their participation costs and increasing the efficiency of their stock investments. Determining the nature of such a policy is well outside the scope of this paper; therefore, rather than choosing a level of effectiveness for this hypothetical intervention, we consider a range of efficiency improvements: 40%, 60%, 80%, and 100%. For each level of improvement, we reduce the stock market participation cost and the risky asset return inefficiency of every agent by that amount . For instance, under an efficiency improvement of 40%, we reduce every household's participation cost by 40% and move every household 40% closer to the stock market's benchmark expected log-return than they are under the baseline estimates.



The compensating variations for both policies are computed at age 21 and from a starting wealth of $20,000. We find the compensating variation for each of our 5,130 agent types and weight them by the number of agents of each type in our simulated population. For both policies and each level of financial inefficiency reduction, we find the median compensating variation at different levels of the EA score. Each point of the graph subtracts the median compensating variation of the benefit reduction policy (at the given inefficiency reduction and EA score) from that of the age-shift policy.

Figure 9: Reductions of Financial Inefficiencies and Welfare Costs.

Figure 9 shows the difference in median compensating variations between the benefit shift and benefit reduction counterfactuals by EA score for the baseline scenario (0% financial inefficiency reduction) along with the four hypothetical levels of inefficiency reduction. As shown previously in Figure 8, under the baseline scenario the distribution of compensating variations are highly similar, with differences in medians generally ranging between $175 and -$200. As stock market inefficiency is reduced, lowering the cost of participation and improving the return on risky assets, the compensating variation in the benefit shift policy becomes significantly larger than in the benefit reduction policy. For instance, in the baseline scenario, households around the center of the EA score distribution (0) were virtually indifferent between each policy, but under perfect efficiency, they would be willing to pay $600 *more* to avoid the benefit shift than they would to avoid the benefit reduction policy.

These results demonstrate that the policies can shift $G \times E$ interactions in complex ways. Not only does the policy environment affect the gradients between observed economic outcomes and genetic endowments, different policies also lead to different distributions of welfare costs for people with different genetic endowments. We highlight that, absent an explicit model that incorporates genetic endowments and endogenous behavior, it is difficult to draw such conclusions.

6. Conclusion

Our analysis demonstrates the important role that economic theory and modeling can play in interpreting the growing literature on gene-by-environment ($G \times E$) interactions. Given that policies affect environments, they can shift $G \times E$ interactions, i.e., how genetic endowments relate to economic outcomes. Our research illustrates that studying $G \times E$ interactions through the lens of an economic model offers two contributions. First, it facilitates *ex-ante* $G \times E$ analysis, by which we mean that we can examine how counterfactual policy environments interact with genes to drive behavior and outcomes. Second, it allows us to examine the welfare consequences of potential policies for people with different genetic endowments. As we have shown, a policy that moderates the gene-wealth gradient may appear to be more equitable and thus more favorable but has similar welfare consequences to a policy that does not moderate the gene-wealth gradient. Future research of this kind can increase our understanding of the specific economic primitives (e.g., expectations, preferences, and constraints) that give rise to genetic gradients in actions and outcomes and thus help to predict how policy changes are likely to affect economic inequality linked to genetic differences. This will require richer models of behavior and, likely, the collection of richer data.

References

B. W. Arold, P. Hufe and M. Stoeckli. 2022. "Genetic Endowments, Educational Outcomes and the Mediating Influence of School Investments." *CESifo Working*

Papers (9841).

O. Attanasio, H. Low and V. Sánchez-Marcos. 2008. "Explaining Changes in Female Labor Supply in a Life-Cycle Model." *American Economic Review* 98 (4):1517–1552.

S. Bagchi. 2015. "Labor Supply and the Optimality of Social Security." *Journal of Economic Dynamics and Control* 58:167–185.

———. 2019. "Differential Mortality and the Progressivity of Social Security." *Journal of Public Economics* 177:104044.

S. Baker, P. Biroli, H. van Kippersluis et al. 2022. "Beyond Barker: Infant Mortality at Birth and Ischaemic Heart Disease in Older Age."

S. H. Barcellos, L. Carvalho and P. Turley. 2021. "The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages, and Later-Life SES."

D. Barth, N. W. Papageorge and K. Thom. 2020. "Genetic Endowments and Wealth Inequality." *Journal of Political Economy* 128 (4):1474–1522.

J. P. Beauchamp, D. Cesarini, M. Johannesson et al. 2011. "Molecular Genetics and Economics." *Journal of Economic Perspectives* 25 (4):57–82.

J. Becker, C. A. P. Burik, G. Goldman et al. 2021. "Resource profile and user guide of the Polygenic Index Repository." *Nature Human Behaviour* 5 (12):1744–1758.

D. W. Belsky, B. W. Domingue, R. Wedow et al. 2018. "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences* 115 (31):E7275–E7284.

D. W. Belsky, T. E. Moffitt, D. L. Corcoran et al. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development." *Psychological Science* 27 (7):957–972.

D. J. Benjamin, D. Cesarini, C. F. Chabris et al. 2012. "The Promises and Pitfalls of Genoeconomics." *Annual Review of Economics* 4 (1):627–662.

B. D. Bernheim, J. Skinner and S. Weinberg. 2001. "What Accounts for the Variation in Retirement Wealth among U.S. Households?" *American Economic Review* 91 (4):832–857.

L. Bierut, P. Biroli, T. J. Galama et al. 2023. "Challenges in studying the interplay of genes and environment: A study of childhood financial distress moderating genetic predisposition for peak smoking." Working paper.

P. Biroli. 2015. "Genetic and Economic Interaction in Health Formation: The Case of Obesity." *Working Paper* :70.

P. Biroli, T. J. Galama, S. von Hinke et al. 2022. "The Economics and Econometrics of Gene-Environment Interplay."

S. E. Black, P. J. Devereux, P. Lundborg et al. 2020. "Poor Little Rich Kids? The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviours." *The Review of Economic Studies* 87 (4):1683–1725.

R. Blundell, M. Costa Dias, C. Meghir et al. 2016. "Female Labor Supply, Human Capital, and Welfare Reform." *Econometrica* 84 (5):1705–1753.

L. E. Calvet, J. Y. Campbell and P. Sodini. 2007. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." *Journal of Political Economy* 115 (5):707–747.

C. D. Carroll. 2002. "Why Do the Rich Save So Much?" In *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, edited by Joel B. Slemrod. Harvard University Press.

S. Catherine. 2021. "Countercyclical Labor Income Risk and Portfolio Choices over the Life Cycle." *The Review of Financial Studies* :hhab136.

D. Cesarini, C. T. Dawes, M. Johannesson et al. 2009. "Genetic Variation in Preferences for Giving and Risk Taking*." *The Quarterly Journal of Economics* 124 (2):809–842.

D. Cesarini, M. Johannesson, P. Lichtenstein et al. 2010. "Genetic Variation in Financial Decision-Making." *The Journal of Finance* 65 (5):1725–1754.

J. Chai, W. Horneff, R. Maurer et al. 2011. "Optimal Portfolio Choice over the Life Cycle with Flexible Work, Endogenous Retirement, and Lifetime Payouts*." *Review of Finance* 15 (4):875–907.

J. F. Cocco, F. Gomes and P. J. Maenhout. 2005. "Consumption and Portfolio Choice over the Life Cycle." *The Review of Financial Studies* 18 (2):491–533.

M. D. Collado, I. Ortuño-Ortín and J. Stuhler. 2023. "Estimating Intergenerational and Assortative Processes in Extended Family Data." *The Review of Economic Studies* 90 (3):1195–1227.

J. C. Conesa and D. Krueger. 1999. "Social Security Reform with Heterogeneous Agents." *Review of Economic Dynamics* 2 (4):757–795.

H. Cronqvist and S. Siegel. 2014. "The Genetics of Investment Biases." *Journal of Financial Economics* 113 (2):215–234.

———. 2015. "The Origins of Savings Behavior." *Journal of Political Economy* 123 (1):123–169.

M. De Nardi, E. French and J. B. Jones. 2010. "Why Do the Elderly Save? The Role of Medical Expenses." *Journal of Political Economy* 118 (1):39–75.

K. E. Dynan, J. Skinner and S. P. Zeldes. 2004. "Do the Rich Save More?" *Journal of Political Economy* 112 (2):397–444.

A. Fagereng, C. Gottlieb and L. Guiso. 2017. "Asset Market Participation and Portfolio Choice over the Life-Cycle." *The Journal of Finance* 72 (2):705–750.

A. Fagereng, L. Guiso, D. Malacrino et al. 2020. "Heterogeneity and Persistence in Returns to Wealth." *Econometrica* 88 (1):115–170.

A. Fagereng, M. Mogstad and M. Rønning. 2021. "Why Do Wealthy Parents Have Wealthy Children?" *Journal of Political Economy* 129 (3):703–756.

J. M. Fletcher and Q. Lu. 2021. "Health policy and genetic endowments: Understanding sources of response to Minimum Legal Drinking Age laws." *Health Economics* 30 (1):194–203.

S. Furuya, J. M. Fletcher, Z. Zhao et al. 2022. "Detecting genetic heterogeneities in response to trauma: The case of 9/11." *SSM - Mental Health* 2.

L. Fuster, A. İmrohoroğlu and S. İmrohoroğlu. 2003. "A Welfare Analysis of Social Security in a Dynastic Framework*." *International Economic Review* 44 (4):1247–1274.

———. 2007. "Elimination of Social Security in a Dynastic Framework." *The Review of Economic Studies* 74 (1):113–145.

H.-M. V. Gaudecker. 2015. "How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice?" *The Journal of Finance* 70 (2):489–507.

G.-L. Gayle, L. Golan and M. A. Soytas. 2022. "What is the source of the intergenerational correlation in earnings?" *Journal of Monetary Economics* 129:24–45.

F. Gomes and A. Michaelides. 2005. "Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence." *The Journal of Finance* 60 (2):869–904.

J.-O. Hairault, F. Langot and T. Sopraseuth. 2008. "Quantifying the Laffer Curve on the Continued Activity Tax in a Dynastic Framework*." *International Economic*

Review 49 (3):755–797.

K. P. Harden. 2021. *The Genetic Lottery: Why DNA Matters for Social Equality*. Princeton University Press, 1st ed.

J. Heckman. 1974. "Life Cycle Consumption and Labor Supply: An Explanation of the Relationship between Income and Consumption Over the Life Cycle." *The American Economic Review* 64 (1):188–194.

W. D. Hill, N. M. Davies, S. J. Ritchie et al. 2019. "Genome-Wide Analysis Identifies Molecular Systems and 149 Genetic Loci Associated with Income." *Nature Communications* 10 (1):5741.

M. Houmark, V. Ronda and M. Rosholm. 2020. "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills." SSRN Scholarly Paper ID 3708642, Social Science Research Network, Rochester, NY.

R. G. Hubbard, J. Skinner and S. P. Zeldes. 1995. "Precautionary Saving and Social Insurance." *Journal of Political Economy* 103 (2):360–399.

S. Imrohoroğlu and S. Kitao. 2012. "Social Security Reforms: Benefit Claiming, Labor Force Participation, and Long-Run Sustainability." *American Economic Journal: Macroeconomics* 4 (3):96–127.

T. Kaiser, A. Lusardi, L. Menkhoff et al. 2022. "Financial Education Affects Financial Knowledge and Downstream Behaviors." *Journal of Financial Economics* 145 (2, Part A):255–272.

T. Kaiser and L. Menkhoff. 2017. "Does Financial Education Impact Financial Literacy and Financial Behavior, and If So, When?" *The World Bank Economic Review* 31 (3):611–630.

R. Karlsson Linnér, P. Biroli, E. Kong et al. 2019. "Genome-Wide Association Analyses of Risk Tolerance and Risky Behaviors in over 1 Million Individuals Identify Hundreds of Loci and Shared Genetic Influences." *Nature Genetics* 51 (2):245–257.

R. Kaygusuz. 2015. "Social Security and Two-Earner Households." *Journal of Economic Dynamics and Control* 59:163–178.

A. Kong, G. Thorleifsson, M. L. Frigge et al. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359 (6374):424–428.

J. J. Lee, R. Wedow, A. Okbay et al. 2018. "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals." *Nature Genetics* 50 (8):1112–1121.

H. Low, C. Meghir and L. Pistaferri. 2010. "Wage Risk and Employment Risk over the Life Cycle." *American Economic Review* 100 (4):1432–1467.

A. Lusardi, P.-C. Michaud and O. S. Mitchell. 2017. "Optimal Financial Knowledge and Wealth Inequality." *Journal of Political Economy* 125 (2):431–477.

A. R. Martin, C. R. Gignoux, R. K. Walters et al. 2017. "Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations." *The American Journal of Human Genetics* 100 (4):635–649.

R. C. Merton. 1969. "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case." *The Review of Economics and Statistics* 51 (3):247–257.

D. Muslimova, H. van Kippersluis, C. A. Rietveld et al. 2020. "Dynamic Complementarity in Skill Production: Evidence From Genetic Endowments and Birth Order." SSRN Scholarly Paper 3748468, Social Science Research Network, Rochester, NY.

A. Okbay, J. P. Beauchamp, M. A. Fontana et al. 2016. "Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment." *Nature* 533 (7604):539–542.

A. Okbay, Y. Wu, N. Wang et al. 2022. "Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals." *Nature Genetics* 54 (4):437–449.

A. Olsen and R. Hudson. 2009. "Social Security Administration's Master Earnings File: Background Information." *Social Security Bulletin* 82 (4).

N. W. Papageorge and K. Thom. 2020. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." *Journal of the European Economic Association* 18 (3):1351–1399.

E. Plug and W. Vijverberg. 2003. "Schooling, Family Background, and Adoption: Is It Nature or Is It Nurture?" *Journal of Political Economy* 111 (3):611–641.

C. A. Rietveld, S. E. Medland, J. Derringer et al. 2013. "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment." *Science* 340 (6139):1467–1471.

V. Ronda, E. Agerbo, D. Bleses et al. 2022. "Family Disadvantage, Gender, and the Returns to Genetic Human Capital*." *The Scandinavian Journal of Economics*

124 (2):550–578.

A. Rustichini, W. G. Iacono, J. J. Lee et al. 2023. "Educational Attainment and Intergenerational Mobility: A Polygenic Score Analysis." *Journal of Political Economy* .

P. A. Samuelson. 1969. "Lifetime Portfolio Selection By Dynamic Stochastic Programming." *The Review of Economics and Statistics* 51 (3):239–246.

J. K. Scholz, A. Seshadri and S. Khitatrakun. 2006. "Are Americans Saving "Optimally" for Retirement?" *Journal of Political Economy* 114 (4):607–643.

Social Security Administration. 2022a. "Proposals to Change Social Security." https://www.ssa.gov/oact/solvency/index.html.

———. 2022b. "Summary of Provisions That Would Change the Social Security Program." Tech. rep., Social Security Administration.

P. Taubman. 1976. "The Determinants of Earnings: Genetics, Family, and Other Environments: A Study of White Male Twins." *The American Economic Review* 66 (5):858–870.

S. Trejo and B. W. Domingue. 2018. "Genetic Nature or Genetic Nurture? Introducing Social Genetic Parameters to Quantify Bias in Polygenic Score Analyses." *Biodemography and Social Biology* 64 (3-4):187–215.

P. M. Visscher, N. R. Wray, Q. Zhang et al. 2017. "10 Years of GWAS Discovery: Biology, Function, and Translation." *The American Journal of Human Genetics* 101 (1):5–22.

A. Vissing-Jorgensen. 2002. "Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures." Working Paper 8884, National Bureau of Economic Research.

A. I. Young, M. L. Frigge, D. F. Gudbjartsson et al. 2018. "Relatedness Disequilibrium Regression Estimates Heritability without Environmental Bias." *Nature Genetics* 50 (9):1304–1310.

ONLINE APPENDIX

A. Genome-Wide Association Studies and EA Score

A.1. Polygenic Scores

In this section, we describe how the EA polygenic score is constructed, what it measures, and some earlier results on its association with socioeconomic outcomes. More detailed accounts are found in past studies from economics. We also refer the reader to Beauchamp et al. (2011), Benjamin et al. (2012), and Visscher et al. (2017) for excellent reviews, upon which we base much of the information presented here.

The human genome consists of approximately three billion pairs of nucleotide molecules, called base pairs, spread out over 23 chromosomes. Four molecules, adenine (A), cytosine (C), guanine (G), and thymine (T), combine to form the two base pair molecules in human DNA: AT or GC. Humans have two copies of each chromosome, one from each parent. Therefore, at a particular location in the genome, an individual may have two copies of AT (AT,AT), one copy of AT and one copy of CG (AT,CG), or no copies of AT and two copies of CG (CG,CG). Sequences of base pairs are called genes and govern bodily function through the synthesis of proteins. At over 99% of locations in the human genome, people have the same base pairs. Locations where there is variation across humans are called *single nucleotide polymorphisms* (henceforth *SNPs*). A "reference allele" is the nucleotide pair that is most common at a given SNP, and an individual can have zero, one, or two copies of the reference allele. For example, at a particular SNP, if the most common base pair is AT, a person's genotype could be (AT, AT), (AT, GC), (GC, AT), or (GC, GC). That person would possess two, one, one, or zero copies of the reference allele, respectively.

Genome-wide association studies (GWAS) are empirical exercises that relate SNP-level data to behaviors and outcomes such as height, body mass, certain diseases, or socioeconomic outcomes and choices. The process typically consists of regressing the outcome of interest onto count variables for the number of reference alleles each individual has at each SNP. Separate regressions are run for each SNP, while controlling for the principal components of the full matrix of SNP count data. The coefficients estimated in GWAS are used to construct *polygenic scores*, linear indices based on the estimated regression coefficients associated with each SNP. These scores are essentially weighted averages of genetic markers, where the weights are the coefficients linking the SNPs to outcomes.[1] Loosely speaking, a higher polygenic score means an individual possesses more of the genetic markers that are correlated with the outcome. The sample used to conduct a GWAS is often referred to as its discovery sample. Polygenic scores can be computed outside of their discovery sample by combining the estimated GWAS coefficients with SNP-level data from a different sample.

The polygenic score that we use in this paper is based on results from Lee et al. (2018), who conducted a GWAS for educational attainment that featured a discovery sample of over 1.1 million people. The score constructed by the authors accounts for 10.6% of the variation in the years of education of HRS respondents of European ancestry. This is a notable achievement because individuals from the HRS were not used in the Lee et al. (2018) discovery sample. We refer to this polygenic score as the *EA3 polygenic score*, since it is the score based on the third major GWAS for education attainment, or simply the *EA score*. While the EA score was constructed to predict years of education, Lee et al. (2018) show that it also has out-of-sample predictive power for related outcomes, such as GPA and cognition. More recent studies demonstrate that the EA score is robustly associated with more complex socioeconomic outcomes such as labor income and wealth at retirement, even after flexibly controlling for educational attainment (Barth, Papageorge, and Thom, 2020; Papageorge and Thom, 2020).

A.2. Limitations

There are various limitations and concerns to using genetic data in social scientific analysis, which the earlier literature has discussed. Studies that established some of the basic empirical relationships on which this paper builds and that which guide our modeling assumptions were careful in addressing these concerns and others. First, it is wrong to conclude that the social outcomes that we study are purely biologically determined. Educational attainment, income, and wealth are all influenced by environmental factors (e.g., childhood economic advantages, parental investments, macroeconomic trends, life events, and social policies), along with genetic endowments and the interaction between the two. The EA score summarizes genetic variations that, on average and given current environments, predict educational attainment and

[1] The regression procedure often makes adjustments to address issues like population stratification, multiple hypothesis testing, and correlation between SNPs. See Benjamin et al. (2012) for details.

other socioeconomic outcomes. Furthermore, despite being able to explain a moderate fraction of the aggregate variation in these outcomes, the EA score remains a poor individual-level predictor (Harden and Koellinger, 2020). For these reasons, we refrain from and caution against interpreting an individual's EA score as their "ability."

Second, with gene-outcome gradients, causality is difficult to ascertain. Unadjusted relationships likely capture factors subsumed into an error term that are related to genetic endowments, including family environments. We note, however, that many relationships between genes and outcomes, including the association between the EA score and education, hold in analyses with family fixed effects, though coefficients are smaller. Incorporating standard childhood SES variables into the analysis generates relationships close to those that rely on within-family variation in the EA score. Thus, many studies, including this one, incorporate such variables (e.g., father's occupation and mother's education). In doing so, we argue that we substantially mitigate the concern that any coefficient estimates using the EA score reflect solely environmental factors correlated with genetic endowments, such as household resources during childhood.

Third, a general limitation of polygenic scores is that they lose predictive power when applied to ethnic groups different from those represented in the GWAS discovery sample (see Martin et al., 2017). In the case of the EA score, the discovery sample used by Lee et al. (2018) was comprised of individuals of European ancestry. This fact limits our ability to extrapolate the mechanisms that we study to people of different ethnic groups.[2]

Fourth, since the EA score is an out-of-sample predictor formed using a series of estimated linear models, it could be subject to measurement error and misspecification relative to a theoretically optimal genetic predictor of educational attainment. This limitation would affect the empirical moments that we match and, by extension, our structural estimates. If this process produced classical measurement error—injecting noise in the reduced form relationships between the polygenic score and our outcomes of interest—we would expect these issues to attenuate the relationships that

[2]Related to this issue, it is common practice to address concerns that genetic relationships reflect population stratification by adjusting for the first 10 principal components of the full matrix of genetic data.

we observe and model, making our estimates act as conservative bounds.

A.3. Social Scientific Research Using the EA Score

Research on molecular genetics and social science outcomes often explores gene-by-environment or $G \times E$ interactions. An important goal for this literature is to understand the extent to which modifiable aspects of the environment can attenuate or intensify genetic influences on important outcomes (Fletcher and Conley, 2013). Various papers find evidence that such aspects of the environment can moderate the relationship between the EA score and completed education. Both Papageorge and Thom (2020) and Ronda et al. (2022) find that the association between the EA score and post-secondary education is larger for people growing up in households with higher socioeconomic status. The EA score is also more strongly associated with educational attainment among first-born individuals, suggesting a possible interaction between genetic endowments and parental investments (Muslimova et al., 2020). Moreover, school characteristics and specific educational reforms have been shown to moderate genetic gradients in educational attainment. For example, Arold, Hufe, and Stoeckli (2022) and Trejo and Domingue (2018) estimate interactions between the EA score and school-level characteristics in predicting educational attainment, while Barcellos, Carvalho, and Turley (2021) find that an increase in the school leaving age in the U.K. reduced the association between the EA score and educational attainment (though not later life economic outcomes). Schmitz and Conley (2017) show that Vietnam-era military conscription had an adverse effect on post-service schooling for veterans with below-average EA scores, but find no effect for those with above-average EA scores. Interactions also exist between the EA score and outcomes other than educational attainment. Barth, Papageorge, and Thom (2020) find that the relationship between the EA score and wealth is significantly reduced among people with defined-benefit pension plans. A larger $G \times E$ literature studies a broad range of health outcomes and tests for interactions between environmental factors and polygenic scores for conditions such as obesity, alcohol consumption, and heart disease (Barcellos, Carvalho, and Turley, 2018; Biroli, 2015; Fletcher and Lu, 2021; Baker et al., 2022).

Disciplines other than economics, such as sociology, psychology, and behavioral genetics, have long explored pathways through which traits and outcomes that economists would relate to human capital are produced and transmitted between generations. Studies in these areas recognize the importance of interactions between genetic endowments and environments for intellectual development, educational attainment, and socioeconomic outcomes (Scarr-Salapatek, 1971; Guo and Stearns, 2002; Turkheimer et al., 2003; Nisbett et al., 2012; Belsky et al., 2016, 2018). These interactions have been studied in different settings, at different stages of life and using different methods and, as shown by Tucker-Drob and Bates (2016), the level of complementarity between genetic endowments and childhood socioeconomic status in the production of academic achievement varies between countries, suggesting that the interaction can change as a function of societal conditions and policies. Therefore, examining these interactions can enhance our understanding of whether social policies reduce or entrench existing inequalities (Harden and Koellinger, 2020).

A natural next step in the use of genetic data in the social sciences is to incorporate the heterogeneity and interactions identified from studying polygenic scores into models of human capital accumulation, life-cycle decisions, and social policies. As Harden and Koellinger (2020) put it, social scientists can use polygenic scores as a type of "tracer dye" to reveal how predispositions, environments, policies, and life events interact in the production of adulthood outcomes, which is consistent with looking more carefully at dynamic life-cycle behavior. Moreover, while earlier work shows relationships between genetic endowments and socioeconomic differences, it cannot tell us how these relationships would change in counterfactual settings, for instance different social policies, labor market conditions, tax systems, or incentives to education. Interactions between genetic data and previous policy changes allow the researcher to conduct ex-post policy evaluations, e.g., to understand how education policy affected people with different EA scores (Barcellos, Carvalho, and Turley, 2018). In contrast, *ex-ante* policy analysis requires an explicit model to simulate counterfactual policies or environments (Wolpin, 2013). Genetic data combined with models of economic behavior permit examination of counterfactual gene-by-environment interactions.

To our knowledge, the only other paper that develops and estimates a structural model to examine genetic endowments related to education attainment is Houmark, Ronda, and Rosholm (2020).[3] The authors incorporate genetic data into a model of

[3] Biroli (2015) uses a structural model to examine genetic markers related to obesity.

Table B1: Non-estimated parameters.

Parameter	Symbol	Value	Parameter	Symbol	Value
Relative risk aversion	ω	1.6 [a]	Yearly discount factor	β	0.96
Consumption's share of utility	γ	0.39 [b]	Minimum and maximum retirement years	$\{\text{Age}_0^R, \text{Age}_f^R\}$	$\{62, 80\}$
Yearly survival probabilities	$\{\delta_t\}_{t=1}^T$	SSA[c]	Minimum and "full" soc. sec. benefit claiming years	$\{\text{Age}_{\min}^{SS}, \text{Age}_{\text{Full}}^{SS}\}$	$\{62, 67\}$
Mean of real S&P500 log-returns	μ^{SP500}	0.0645 [d]	Income tax rate (soc. sec.)	τ^{FICA}	0.06
Std. Dev. of real S&P500 log-returns	σ^{SP500}	0.1649 [d]	Income tax rate	τ^W	0.24
Yearly risk-free return	R	1.015	Capital gains tax rate	τ^c	0.20
			S.S. benefits tax rate	τ^s	0.15

[a] Lusardi, Michaud, and Mitchell (2017), [b] Imrohoroğlu and Kitao (2012), [c] SSA Life Tables for males born in 1940, [d] historical calculation based on years 1936-2018 from the accompanying data file to Chapter 26 of Shiller (1990).

skill formation and parental investments. The structure in their model means they can address questions about counterfactual policies, e.g., they can simulate the relationship between the EA score and children's skills in a counterfactual environment where parental investments are held fixed across the population. In our case, we develop a model of life-cycle behavior where choices are an explicit result of dynamic utility maximization under uncertainty. Choices are optimal reactions to endowments (including genetic endowments) and the environment, and they endogenously adjust to changes in policies. The model also allows us to measure welfare so that we can understand how counterfactual policies not only affect optimal choices and outcomes but also well-being for people with different genetic endowments. As we show, this is important since policies that reduce inequality on some outcomes for people with different EA scores can increase inequality in terms of lifetime utility.

B. Estimation details

We split our model's parameters in three different groups. Table B1 presents the values of the first set, which we take from historical data, administrative sources, or past studies. We estimate the rest of the parameters in two steps as described in the main text. The following subsections expand on the details of both steps.

Table B2: Income, Inheritance, and Defined-Benefit Pension Estimates.

	[1] Log Real Earnings Linear, Indiv. F.E.	[2] P^I Probit	[3] ln(C.Inher), OLS	[4] ln D.B. Pension OLS
Const.	8.6266	−6.03	11.70	−14.10
Age	0.1092	0.13	−0.04	
Age^2/100	−0.1250	−0.10	0.03	
Coll	−0.3027	0.13	0.53	0.69
EA	−0.0405	0.02	0.15	
EA Quint. 2				0.11
EA Quint. 3				0.23
EA Quint. 4				0.39
EA Quint. 5				0.20
SES	0.0057			
SES.Miss.	−0.0051			
EA × SES.Miss.	−0.0199			
Age × Coll	0.0099			
Age × EA	0.0022			
Age × SES	0.0016			
Age × SES.Miss.	0.0085			
Year	0.0075			
BY				0.01
DB.Ever	0.1342			
DB.Always	0.0411			
DB.Miss.	0.0864			
Unemp	−0.0648			
Unemp^2	0.0033			
Princ. Comp. Gen.	✓	✓	✓	
Num. obs.	31883	49000	1056	1,672
$\sigma(\mathcal{Z}_i^w)$	0.3740			
$\sigma(\epsilon_{i,t}^w)$	0.3867			
Sigma OLS	0.4947		1.53	

We control for the 10 first principal components of genetic data as released in the SSGAC file for the EA polygenic score (Lee et al., 2018).

B.1. First Step Estimation Details

In the first step, we directly estimate the specifications of wages (Equation 6), inheritances, and defined benefit pension flows on the HRS sample data.

a) Income: Pre-tax labor income is:

$$\ln \tilde{W}_{i,t} = f(\text{Age}_{i,t}, \text{EA}_i, \text{Coll}_i, \text{SES}_i, \text{DB}_i, \text{Year}_t, \text{Unemp}_t) + \mathcal{Z}_i^w + \epsilon_{i,t}^w.$$

We let $f(\cdot)$ be a linear function of its arguments and some of their interactions. We estimate this equation with fixed-effects regressions using Social Security Administration earnings records that have been linked to our HRS sample. The sample for this regression consists of all person-year observations for individuals in our main wealth sample with non-missing SSA earnings data. Our dependent variable is the log of real earnings, taken from the SSA Summary earnings files. We adjust these amounts to replace top-coded values with estimates of the expected value of top-coded amounts as in Barth, Papageorge, and Thom (2020). We consider observations of non-retired individuals between the ages 22-65. We also restrict our sample to person-year observations with earnings of at least $10,000 in real income (with 2010 as the base monetary year). The remaining sample consists of $31,883$ person-year observations from 869 unique individuals.

The first column of Table B2 presents our estimates of the income process. SES.Miss is an indicator for individuals for whom the childhood socioeconomic status index is missing. DB.Ever is an indicator for whether the individual was ever observed receiving a defined benefit pension when retired. DB.Always is an indicator for whether the individual was always observed receiving a defined benefit pension when retired. DB.Miss is an indicator for individuals for whom the information about defined benefit pensions is missing.

b) Inheritances: We estimate both the probability of receiving an inheritance, P^I, and the value of an inheritance conditional on its receipt, C.Inher, as functions of a person's age, education, and EA score. Table B2 presents our estimates. The second column models the probability of receiving an inheritance using a probit model. The third column presents OLS estimates for a log-linear model of inheritances conditional on receipt. The samples for each of the regressions are the following. First, using retrospective survey items about the timing of receiving an inheritance, we construct a panel of $49,000$ household-year observations for the 870 households in our main wealth sample. Each household-year corresponds to a person-year for the male household member when he was between the ages of 22 and 80. We use this sample to estimate the models of the probability of receiving an inheritance, P^I. For the value of inheritances conditional on reception, we use the sub-sample of $1,056$ household-year observations for 575 households corresponding to periods when they received a non-zero inheritance.

c) Defined Benefit Pension Flows: The fourth column of Table B2 presents our estimates for defined-benefit pension-flows. The sample consists of 1,672 households receiving defined benefit pension income from Barth, Papageorge, and Thom (2020).

B.2. Second Step MSM Estimation Details

We estimate the rest of the parameters internally, targeting moments of interest. Specifically, define $\mathcal{M}$ as the vector of empirical sample moments we target in the estimation. The MSM methodology begins by generating counterparts to $\mathcal{M}$ from a simulated sample, generated conditional on a candidate set of parameters $\widetilde{\theta}$. To generate the synthetic sample, we match the sample data's joint distributions of birth years, college completion, EA polygenic scores, and defined benefit pension plan participation. To do so, we discretize the EA score by its deciles and group individuals into five-year birth cohorts from 1915 to 1960. This gives us 400 possible combinations of observable characteristics: ten EA score deciles, times ten birth cohorts, times two educational attainment levels (a completed college degree or not), times two possible pension possibilities (participating in a defined benefit plan or not). Only 190 out of the 400 combinations are populated in the sample data. Denote by N_q the number of observations in each of the $q = 1, 2, ..., 190$ populated combinations. We expand the synthetic sample by multiplying N_q by 10 for each q. Note that this expands the synthetic sample by a factor of 10 but preserves the relative distribution of characteristics.

We then introduce unobserved heterogeneity. Given a candidate set of parameters and an agent's EA score, Equations (2) and $\text{SES}_i = \phi EA_i + \varepsilon_i$ imply:

$$\vec{\zeta}_i | EA_i \sim \mathcal{N}(z\phi EA_i, \sigma^2_{\text{SES}} zz' + \Sigma_{\mathcal{Z}}). \tag{1}$$

We discretize the distribution in Equation (1) using 27 equiprobable points for $\vec{\zeta}_i$ (three equiprobable values for each of the three sources of unobserved heterogeneity for a total of 27 possible triplets). We set the socioeconomic status of simulated agents to $E[\text{SES}_i|\text{EA}_i, \vec{\zeta}_i]$.[4] Lastly, we simulate 27 agents, one for each possible draw of unobserved heterogeneity, for each of the $N_q \times 10$ agents that populate the 190 populated bins. That is, each of the 190 non-empty combinations of observable characteristics in the HRS sample will contain 27 different simulated agents, each with

[4]With our setup, $E[\text{SES}_i|EA_i, \vec{\zeta}_i] = \phi EA_i + (\sigma^2_{\text{SES}} z')(\sigma^2_{\text{SES}} zz' + \Sigma_{\mathcal{Z}})^{-1}(\vec{\zeta}_i - z\phi EA_i)$.

different unobserved-heterogeneity draws, replicated $N_q \times 10$ times. Expanding the 190 groups of observable characteristics with the 27 EA-specific draws of unobserved heterogeneity leaves us with 5,130 *ex-ante* types of agents. These types constitute the full structure of *ex-ante* heterogeneity.

We then solve the life-cycle problem by backward induction for the 5,130 types, which delivers policy functions for choice variables conditional on states. Appendix E presents details on how we solve the model. We then simulate the lives of our entire population of agents. We generate draws of lifetime labor income following Equation (6) and mortality draws from the survival rates $\{\delta\}_{t=21}^{90}$. For stock-market returns R_t^{SP500}, we take the realized annual return of the S&P 500 index for the relevant year, which is determined by each agent's birth-year and age. This produces a random sample of synthetic life-cycles, and we calculate counterparts of our targeted moments on the simulated data. Note that this entire procedure is conditional on *a given set of parameters* $\widetilde{\theta}$. Denote by $\widehat{\mathcal{M}}(\tilde{\theta})$ the simulated moments conditional on $\widetilde{\theta}$.

We define the loss function for a set of parameters $\tilde{\Theta}$ given K empirical moments as: $L(\tilde{\theta}) = \sum_{k=1}^{K} \left(\frac{\widehat{\mathcal{M}}(\tilde{\theta})_k - \mathcal{M}_k}{\mathcal{M}_k} \right)^2$. We estimate the model by minimizing the loss function, $\hat{\theta} = \arg\min_\theta L(\theta)$.[5]

B.2.1. Empirical Moments to Match

Our internal estimation routine targets the following empirical moments:

a) Wealth distribution: The mean and the $10th$, $25th$, $50th$, $75th$, and $90th$ percentiles of the wealth distribution for people in the age range $60 - 70$.

b) Wealth regression coefficients: A subset of the coefficients from a regression of log-wealth on the EA score EA_i, our index of childhood socioeconomic status SES_i, a binary indicator of college completion Coll_i, the log "earned income" $\ln \text{Earned Income}_i$ which we define as the sum of labor income earned between the ages of 30 and 60, a binary indicator of stock ownership $\text{Stocks}_{i,t}$, and age. We only use observations on people aged 60 to 70 that report strictly positive wealth to estimate the auxiliary model. We target the coefficients on Coll_i, EA_i, SES_i, and $\text{Stocks}_{i,t}$.

[5] We solve the minimization problem using the Nelder-Mead algorithm with multiple starting points. We start by evaluating the objective function in 20,000 initial points and then use the Nelder-Mead algorithm starting at the 25 points with the smallest loss functions. We take the result with the lowest associated loss as our estimate.

c) Stock ownership rate: The fraction of individual-year observations in our sample aged 60 to 70 that report owning any stocks ($\text{Stocks}_{i,t} = 1$).

d) Stock ownership regression: A subset of the coefficients from a regression of the stock ownership indicator $\text{Stocks}_{i,t}$ on the EA score EA_i, our index of childhood socioeconomic status SES_i, a binary indicator of college completion Coll_i, the log of earned income $\ln(\text{Earned Income}_i)$, and age. We include only people aged 60 to 70 that report strictly positive wealth and do not include the coefficient on age as a moment to match.

e) Retirement rates: The fraction of people in each of the following age brackets who are retired: $[50, 62]$, $[63, 67]$, $[68, 72]$, and ≥ 73.

f) Retirement regression: A subset of the coefficients from a regression of an indicator for whether an individual is retired $\text{Retired}_{i,t}$ on the EA score EA_i, our index of childhood socioeconomic status SES_i, a binary indicator of college completion Coll_i, and age. We consider people aged 61 to 75 and only use the coefficients on Coll_i and EA_i as moments in the estimation.

B.2.2. Identification

In this section we briefly discuss how the variation in the empirical moments identifies the set of estimated parameters Θ. In general, the elements of Θ are jointly determined by the entire vector of moments, and a specific parameter is not identified solely by one specific moment. We do, however, select moments intentionally based on how their variability informs the model parameters.

First, consider the parameters associated with stock market participation costs (f_0, f_{Coll}, f_{EA}, $\sigma(\mathcal{Z}^F)$) and stock returns (r_0, r_{Coll}, r_{EA}, $\sigma(\mathcal{Z}^r)$). Intuitively, the joint distribution of income, wealth, and stock market participation identifies the cost of stock market participation and returns on risky assets. Note first that the income process is estimated directly, outside the MSM procedure. Then, holding risky asset returns fixed, the association between stock market participation and income identifies the distribution of participation costs, since higher average costs will predict lower participation at lower levels of income, and a larger variance in unobserved heterogeneity will imply a flatter relationship between income and stock market participation. Similarly, holding income and participation costs fixed, the distribution of wealth determines the returns on risky assets, since in the model the growth rate of savings is determined entirely from the return on assets.

The coefficient on SES in the wealth regression identifies z_R, the part of unobserved heterogeneity in stock returns that is determined by SES. Note that the coefficient estimate is not a direct estimate of z_R; z_R is the unobserved heterogeneity in *returns* that are linked to SES, not wealth. However, the parameter z_R will affect the association between SES and wealth, and therefore the empirical association between SES and wealth implicitly identifies the association between returns and SES. Similarly, the coefficient on SES in the stock market participation regression identifies z_F, the parameter that determines the effect of SES on unobserved stock market participation cost heterogeneity.

In addition to the stock market parameters, the wealth distribution also identifies the two parameters governing the bequest motive, θ and κ. Without a bequest motive, households would spend down their wealth aggressively as they aged and the wealth distribution would be less skewed at the top. Because there are more wealth distribution moments than are needed to identify the coefficients related to stock market returns, the additional information contained in the distribution of wealth identifies the parameters governing households bequest incentives.[6]

Next, consider coefficients on the disutility of labor (d_0, d_{Coll}, d_{EA}, d_{Age}). Conditional on income and wealth, the retirement decision is influenced by how (un)enjoyable is working. The rates of retirement at each of the four age ranges therefore identifies the distribution of labor utility costs.

Finally, the distributions of stock market participation costs, stock market returns, and disutility of labor are influenced by covariates, in particular the EA score and education. The coefficients from regressions of wealth, stock market participation, and retirement on the EA score and education constitute moments that identify how these distributions shift with these controls.

B.2.3. Additional Possible Mechanisms

Our empirical model allows for the PGS to affect lifetime wealth accumulation through multiple channels: the income process, the distribution of returns on invested

[6] We have experimented with alternative models that omit the bequest motive. The key difference is that we fail to match upper-tail wealth, but key qualitative results are unchanged, i.e., our specification of the bequest motive does not drive our main results.

wealth, the fixed cost of stock market participation, the disutility of labor, and the inheritance process. We focus on these mechanisms because the empirical results in Barth, Papageorge, and Thom (2020) highlighted these mechanisms as important in accounting for the PGS-wealth gradient. Here we consider five other possible channels through which variation in the PGS could be linked to changes in wealth.

a) Fertility The PGS could affect wealth through an effect on fertility. *Ceteris paribus*, we would expect more children to reduce parental wealth in retirement given the expenses of raising kids. Here we construct a measure of the number of kids, which is the sum of the maximum numbers of own children, step children, and other children associated with the male household head from the RAND HRS Family Data File 2018 (V2) for all years.

b) Marital History Our sample focuses on intact couples in order to avoid dealing with issues of wealth transfer during divorce. However, it could be the case that the PGS is associated with past divorces, which could reduce wealth if the end of previous marriages resulted in a reduction in wealth. We account for this by measuring the total number of marriages the male household member had as recorded in the RAND HRS Family Data File 2018 (V2)

c) Mortality Expectations Individuals may save more if they expect to live longer. The HRS has repeatedly asked individuals to assign probabilities to the event of surviving to age 75 in the Expectations section of the survey. For each individual, we record their earliest response to this item, as recorded in the RAND HRS Longitudinal File 2020.

d) Risk Tolerance The PGS could be associated with wealth if it operates through risk preferences (or how individuals process trade-offs in risky situations). Since the 2014 wave, the Leave Behind section of the HRS asks respondents to think about "How willing are you to take risks in financial matters" and respond on a 0 to 10 scale, where 0 indicates "unwilling to take any risks" and 10 indicates "fully prepared to take risks." We record the maximum value of this variable as our measure of risk tolerance. (The results below are nearly identical if we use the minimum value or the mean value for those individuals with multiple measures).

e) Business Ownership The PGS could also operate through entrepreneurship and business ownership. To measure this, we create a dummy variable for whether the household is ever observed owning a business across the HRS waves 1992-2010.

As described in the previous subsection, we identify the effects of variation in the PGS on stock market returns and the cost of stock market participation by matching coefficients from descriptive regressions that predict log wealth and a dummy for any stock holding as a function of the PGS and other key variables in the theory. In particular, the coefficients on the PGS in predicting each of these key outcomes provides critical identifying information. In Table B3, we assess whether controlling for the five additional mechanisms listed above would substantially alter the magnitude of the coefficients on the PGS in these regressions. Panel A presents regression results with log Wealth as the dependent variable, and Panel B presents regression results with a dummy variable for holding any stocks as the dependent variable. All regressions contain the full set of explanatory variables and controls included in the analogous (full) specifications in Table 2. Column (1) of each Panel recreates the main regression specification whose coefficients we match in our estimator. Column (2) in each panel repeats the specification in Column (1), but restricts the sample to the subsample with non-missing values of the five additional mechanism variables described above. Restricting the sample significantly reduces the sample size to 1574, largely because of missingness in the measures of risk aversion. The coefficients on the PGS in each panel are slightly attenuated compared to the Column (1) results. Columns (3) - (7) present results when the specification in Column (1) is modified to include controls for the different mechanisms. Each of these columns adds only one more control beyond the baseline specification. In the case of the wealth regressions, all of these extra mechanism variables are significantly associated with log wealth, in theoretically predicted directions. Adding these controls produces estimates of the coefficient on the PGS that range from 0.130 (controlling for risk aversion) to 0.160 (controlling for longevity expectations). In all cases the coefficient remains highly statistically significant. In Column (8), we add all of the extra mechanism controls simultaneously. This produces a coefficient of 0.126 on the PGS, which is again remains statistically significant. Columns (3) - (8) in Panel B of Table B3 perform the same exercise for our regression with any stockholding as the outcome variable. Across all specifications, we continue to estimate a statistically significant coefficient on the EA Score that is comparable to the baseline estimates of 0.047 (using the restricted sample in Column 2).

Overall, the results presented here suggest that controls measuring the five additional channels listed above only slightly moderate our estimate associations between the EA Score, log wealth, and owning stocks. Of course, our measures here are imperfect and measured with error. However, the fact that all of them show up as individually significant predictors of wealth suggest that they do capture important signals about these underlying channels. Controlling for all of these simultaneously attenuates the estimated coefficients on the EA Score moderately (about 14 percent for log wealth and 13 percent for any stocks). These results suggest that the additional mechanisms explored here could be relevant but are unlikely to account for the majority of the PGS - wealth relationship we study. Given the computational expense of adding additional mechanisms and channels, we have tailored our model to focus on earnings, stock market returns, labor supply, and the inheritance flows.

C. Computing Social Security Benefits

This section describes the procedure we follow for calculating the Social Security benefit schedules in our model. We aim to replicate the real methodology used by the Social Security Administration (Social Security Administration, 2019, Section 2) with one simplification that eases the computational costs of solving our model.

C.1. The Average Indexed Monthly Earnings AIME

The basis of Social Security benefits is a worker's average earnings over his 35 highest-earning years. Tracking this object precisely would require us to incorporate additional state variables to our model. Instead, we make the simplification that agent i's benefits are calculated based on the wage trajectory that someone with i's characteristics would *expect*. Formally, following Equation 6, we approximate agent i's expected pre-tax wages as:

$$\tilde{W}^E_{i,t} = \exp\left\{ f(\text{Age}_{i,t}, \text{EA}_i, \text{Coll}_i, \text{SES}_i, \text{DB}_i, \text{Year}_t, \bar{\text{Unemp}}) + \mathcal{Z}^w_i + \frac{\sigma^2(\epsilon^w)}{2} \right\},$$

where $\bar{\text{Unemp}}$ is the average historical unemployment rate. The computational convenience comes from the fact that $\tilde{W}^E_{i,t}$ is the same for every agent with the same *ex-ante* characteristics as i and therefore tracking it requires no additional state variables.

For some given retirement age $A \in [\text{Age}^R_0, \text{Age}^R_f]$, denote with $\tilde{W}^*_{i,A}$ the 35 greatest

Table B3: Descriptive Regressions: Additional Possible Mechanisms.

Panel A: log Wealth	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
EA Score	0.152***	0.147***	0.155***	0.157***	0.160***	0.130***	0.135***	0.126***
	(0.045)	(0.050)	(0.045)	(0.044)	(0.045)	(0.050)	(0.043)	(0.048)
Num. Kids			-0.047***					-0.040*
			(0.016)					(0.023)
Num. Marriages				-0.194***				0.021
				(0.070)				(0.077)
Prob. Live 75					0.003**			0.003*
					(0.001)			(0.002)
Risk Tolerance						0.072***		0.062***
						(0.017)		(0.017)
Has Business							0.620***	0.464***
							(0.085)	(0.085)
N	2259	1574	2259	2255	2209	1598	2259	1574
R^2	0.380	0.361	0.386	0.385	0.384	0.377	0.419	0.412
Panel B: Any Stocks	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
EA Score	0.049***	0.047***	0.049***	0.050***	0.051***	0.042**	0.048***	0.041***
	(0.014)	(0.017)	(0.014)	(0.014)	(0.014)	(0.016)	(0.014)	(0.016)
Num. Kids			-0.004					0.002
			(0.005)					(0.006)
Num. Marriages				-0.036				-0.041
				(0.022)				(0.030)
Prob. Live 75					0.001***			0.001
					(0.001)			(0.001)
Risk Tolerance						0.024***		0.023***
						(0.007)		(0.007)
Has Business							0.038	0.025
							(0.027)	(0.030)
N	2259	1574	2259	2255	2209	1598	2259	1574
R^2	0.186	0.160	0.187	0.188	0.191	0.178	0.188	0.184

This table reports results from models predicting log Wealth (Panel A) and a dummy variable for any stocks (Panel B). All regressions include the following variables: the first 10 principal components of the genetic data, the family SES score, an indicator for a missing SES score, an indicator for having a college degree, log of Prime Income, an indicator for holding any stocks (in the wealth specification), and interactions between the indicator for a missing SES score and the following: the EA Score, College, log Prime Income, or Any Stocks if they are present in the specification.

elements of the set: $\left\{\min\left[\tilde{W}^E_{i,t}, \bar{T}(\text{Year}_t)\right]\right\}_{t=t(21)}^{t(A)-1}$, where $t(a)$ is an auxiliary function indicating the simulation period at which the agent reaches age a. The set represents the 35 highest-earning years of i up to age A capped at the time-varying maximum taxable earnings. For each retirement age, we find the average indexed monthly earnings as $\text{AIME}_i^A = \frac{1}{12} \times \frac{1}{35} \sum_{W \in \tilde{W}^*_{A,i}} W$.

C.2. Primary Insurance Amount and Adjustments

The primary insurance amount (PIA) is the monthly benefit that a person retiring at the *full retirement age* of 67 would receive. It is a piece-wise linear function of the AIME that has decreasing replacement rates for different income brackets. We use the bracket limits and replacement rates defined by the SSA,

$$\mathrm{PIA}(\mathrm{AIME}_i^A) = \left[\mathrm{AIME}_i^A\right]_0^{895} \times 0.9 + \left[\mathrm{AIME}_i^A\right]_{859}^{5397} \times 0.32 + \left[\mathrm{AIME}_i^A\right]_{5397}^{\infty} \times 0.15$$

where $[x]_a^b \equiv \max\{\min\{x,b\} - a, 0\}$.

The final step is to adjust the PIA depending on whether a person retires before or after the full retirement age. Benefits are reduced by 5/9 percentage points per month of early retirement up to 36 months and by 5/12 percentage points for each additional month over 36. On the other hand, for agents retiring after the full retirement age, benefits increase by 16/24 percentage points for each month not receiving benefits up to age 70, when benefits stop increasing. Formally, for an age of retirement A, defining $H = A - 67$, we obtain yearly benefits as:

$$\mathrm{SSB}_i^A = \begin{cases} 12 \times \mathrm{PIA}_i^A \times \left(1 - \frac{\frac{5}{9}\min\{12|H|,36\} + \frac{5}{12}\max\{12|H|-36,0\}}{100}\right), & \text{If } H < 0 \\ 12 \times \mathrm{PIA}_i^A \times \left(1 + \frac{\frac{16}{24}\min\{H, 3\times 12\}}{100}\right), & \text{If } H \geq 0 \end{cases}. \tag{2}$$

D. Recursive Representation of the Model

This section presents all the dynamic optimization problems that our model's agents can face depending on their age and retirement status. We drop the individual sub-indices i in this section.

a) Between ages 21 and $\mathrm{Age}_0^R - 2$, the agent has not retired and does not have the option to retire next period:

$$V_t(A_t,\emptyset) = \max_{C_t,\phi_t} u(C_t, 1) + \beta\delta_t \mathbb{E}_t\left[V_{t+1}(A_{t+1},\emptyset)\right] + \not{\delta}_t \varphi(S_t)$$

$$\text{Subject to:} \quad 0 \leq C_t, \quad 0 \leq S_t, \quad 0 \leq \phi_t \leq 1$$

$$S_t = A_t - C_t - \mathrm{F}_i \times \mathbf{1}\left[\phi_t > 0\right],$$

$$A_{t+1} = \left\{(1-\tau^c)\left[\phi_t \tilde{R}_{t+1} + (1-\phi_t)R\right] + \tau^c\right\} \times S_t + \tau_t(\tilde{W}_{t+1}) + \mathrm{Inher}_{t+1}.$$

b) Non-retired agents whose age satisfies $\mathrm{Age}_t \in [\mathrm{Age}_0^R - 1, \mathrm{Age}_f^R - 2]$ decide whether to retire or not (with the choice becoming effective next period):

$$
\begin{aligned}
V_t(A_t,\emptyset) = &\max_{C_t,\phi_t,\mathtt{Ret.Age}_{t+1}} u(C_t,1) + \beta\delta_t\mathbb{E}_t\left[V_{t+1}(A_{t+1},\mathtt{Ret.Age}_{t+1})\right] + \not{\delta}_t\varphi(S_t) \\
&\text{Subject to:} \quad 0 \le C_t, \quad 0 \le S_t, \quad 0 \le \phi_t \le 1, \quad \mathtt{Ret.Age}_{t+1} \in \{\emptyset, \mathrm{Age}_t + 1\} \\
&S_t = A_t - C_t - \mathrm{F} \times \mathbf{1}\left[\phi_t > 0\right], \\
A_{t+1} = &\left\{(1-\tau^c)\left[\phi_t\tilde{R}_{t+1} + (1-\phi_t)R\right] + \tau^c\right\} \times S_t + \mathrm{Inher}_{t+1} + \\
&\mathbf{1}[\mathtt{Ret.Age}_{t+1} = \emptyset] \times \tau_t(\tilde{W}_{t+1}) + \mathbf{1}[\mathtt{Ret.Age}_{t+1} \neq \emptyset] \times (1-\tau^s) \times \mathrm{DBf} + \\
&\mathbf{1}[\mathtt{Ret.Age}_{t+1} \neq \emptyset, \mathrm{Age}_{t+1} \ge \mathrm{Age}^{SS}_{\min}] \times (1-\tau^s) \times \mathrm{SSB}(\mathtt{Ret.Age}_{t+1}).
\end{aligned}
$$

c) At $\mathrm{Age}_t = \mathrm{Age}^R_f - 1$ an agent who has not retired is forced to retire:

$$
\begin{aligned}
&V_t(A_t,\emptyset) = \max_{C_t,\phi_t} u(C_t,1) + \beta\delta_t\mathbb{E}_t\left[V_{t+1}(A_{t+1},\mathtt{Ret.Age}_{t+1})\right] + \not{\delta}_t\varphi(S_t) \\
&\text{Subject to:} \quad 0 \le C_t, \quad 0 \le S_t, \quad 0 \le \phi_t \le 1, \quad \mathtt{Ret.Age}_{t+1} = \mathrm{Age}^R_f \\
&S_t = A_t - C_t - \mathrm{F} \times \mathbf{1}\left[\phi_t > 0\right], \\
&A_{t+1} = \left\{(1-\tau^c)\left[\phi_t\tilde{R}_{t+1} + (1-\phi_t)R\right] + \tau^c\right\} \times S_t + (1-\tau^s) \times \mathrm{DBf} + \\
&\qquad \mathbf{1}[\mathrm{Age}_{t+1} \ge \mathrm{Age}^{SS}_{\min}] \times (1-\tau^s) \times \mathrm{SSB}(\mathtt{Ret.Age}_{t+1}) + \mathrm{Inher}_{t+1}.
\end{aligned}
$$

d) For an agent who has already retired and did so at age n ($\mathtt{Ret.Age}_t = n$):

$$
\begin{aligned}
&V_t(A_t,n) = \max_{C_t,\phi_t} u(C_t,0) + \beta\delta_t\mathbb{E}_t\left[V_{t+1}(A_{t+1},n)\right] + \not{\delta}_t\varphi(S_t) \\
&\qquad \text{Subject to:} \quad 0 \le C_t, \quad 0 \le S_t, \quad 0 \le \phi_t \le 1 \\
&\quad S_t = A_t - C_t - \mathrm{F} \times \mathbf{1}\left[\phi_t > 0\right], \\
&A_{t+1} = \left\{(1-\tau^c)\left[\phi_t\tilde{R}_{t+1} + (1-\phi_t)R\right] + \tau^c\right\} \times S_t + (1-\tau^s) \times \mathrm{DBf} + \\
&\qquad \mathbf{1}[\mathrm{Age}_{t+1} \ge \mathrm{Age}^{SS}_{\min}] \times (1-\tau^s) \times \mathrm{SSB}(n) + \mathrm{Inher}_{t+1}.
\end{aligned}
$$

e) In the terminal period, the agent simply allocates his assets between consumption and bequests. He won't be working, provided that the maximum age of retirement is lower than the oldest possible age. His problem is:

$$
V_T(A_T,\mathtt{Ret.Age}_T) = \max_{C_T \ge 0} u(C_T,0) + \varphi(A_T - C_T). \tag{3}
$$

E. Solution of the Life-Cycle Model

We solve the model using value-function iteration for each of the $5,130$ agent types. This section describes how we solve each agent's dynamic problem. We drop the i sub-indices for compactness.

E.1. Grids and Discretizations

Agents have two continuous choices: consumption C and the risky-asset portfolio share ϕ. We discretize C expressing it as a fraction of the assets available for consumption, starting at 0.005 until 1.000 in increments of 0.005. We discretize the risky asset portfolio share using 11 points, from 0.0 to 1.0 in increments of 0.1. Finally, we construct a grid that we use to represent discretizations of various versions of wealth (e.g., start-of period, savings) and generically refer to it as the wealth grid. The wealth grid has 51 points spanning $[10^2, 10^7]$ USD and is denser at lower values.

We also discretize the two normal random variables over which agents form expectations: the income shock ϵ^w and the market log-return $\ln R^{\text{SP500}}$. We use equiprobable grids with 9 points for each.

E.2. Solving the Terminal Period

An agent that reaches the terminal age of 90 faces no portfolio decision, he simply allocates his wealth between consumption and bequests. Since 90 is also larger than the maximum age of retirement, we know the agent will be retired. Therefore, the agent's value function is given by Equation 3. We solve this problem for every A_{90} in our wealth grid with consumption taking the values of our discretized proportional grid. The solution is the same for every retirement status $\texttt{Ret.Age}_{90}$. We use the solutions to construct linear interpolators for $V_{90}(\cdot, \texttt{Ret.Age}_{90})$ and move on to non-terminal periods.

E.3. Solving Non-Terminal Periods

Given value function $V_{t+1}(\cdot,\cdot)$, we must solve for $V_t(\cdot,\cdot)$. We start by defining the function $\texttt{Emax}_{t+1}(\cdot,\cdot,\cdot)$ as:

$$
\begin{aligned}
&\texttt{Emax}_t(S_t, \phi_t, \texttt{Ret.Age}_{t+1}) = E_t\left[V_{t+1}(A_{t+1}, \texttt{Ret.Age}_{t+1})\right] \\
&\text{where} \quad A_{t+1} = X_{t+1} + Y_{t+1} \\
&X_{t+1} = \left\{(1-\tau^c)\left[\phi_t \tilde{R}_{t+1} + (1-\phi_t)R\right] + \tau^c\right\} \times S_t \\
&Y_{t+1} = \mathbf{1}[\texttt{Ret.Age}_{t+1} = \emptyset] \times \tau_t(\tilde{W}_{t+1}) + \mathbf{1}[\texttt{Ret.Age}_{t+1} \neq \emptyset] \times (1-\tau^s) \times \text{DBf}+ \\
&\mathbf{1}[\texttt{Ret.Age}_{t+1} \neq \emptyset, \text{Age}_{t+1} \geq \text{Age}^{SS}_{\text{min}}] \times (1-\tau^s) \times \text{SSB}(\texttt{Ret.Age}_{t+1}) + \text{Inher}_{t+1}.
\end{aligned}
\tag{4}
$$

$\texttt{Emax}$ represents the continuation value from each of the possible states in which the agent can end period t. In Equation 4, we have separated A_{t+1} into wealth coming from past savings X_{t+1} and current income Y_{t+1}. These two components are independent from the point of view of period t and we take advantage of this fact.

We define $\mathcal{Q}_t(X_{t+1}, \texttt{Ret.Age}_{t+1}) \equiv E[V_{t+1}(X_{t+1}+Y_{t+1}, \texttt{Ret.Age}_{t+1})|X_t]$, where the expectation is being taken over Inher_{t+1} and income shock ϵ^w_{t+1}. For every feasible value of $\texttt{Ret.Age}_{t+1}$, we evaluate this function with X_{t+1} on our wealth grid and use the results to construct linear interpolators of $\mathcal{Q}_{t+1}(\cdot, \texttt{Ret.Age}_{t+1})$.

Then, for every feasible $\texttt{Ret.Age}_{t+1}$ and every ϕ_t, we evaluate:

$$\texttt{Emax}_{t+1}(S_t, \phi_t, \texttt{Ret.Age}_{t+1}) = E[\mathcal{Q}_{t+1}(X_{t+1}, \texttt{Ret.Age}_{t+1})|S_t, \phi_t]$$

with S_t taking all the values in our wealth grid and the expectation being taken over market returns only. We use the results to construct linear interpolators of $\texttt{Emax}_{t+1}(\cdot, \phi_t, \texttt{Ret.Age}_{t+1})$.

We now turn to finding the agent's optimal choices. It is useful to note at this point that we can use Equation 4 to re-express the agent's problem as:

$$V_t(A_t, \texttt{Ret.Age}_t) = \max_{C_t, \phi_t, \texttt{Ret.Age}_{t+1}} u(C_t, \ell) + \beta\delta_t \texttt{Emax}_{t+1}(S_t, \phi_t, \texttt{Ret.Age}_{t+1}) + \not{\delta}\varphi(S_t)$$
$$\text{Subject to:} \quad 0 \leq C_t, \quad 0 \leq S_t, \quad 0 \leq \phi_t \leq 1, \quad S_t = A_t - C_t - \mathrm{F} \times \mathbf{1}[\phi_t > 0],$$

with the feasible options for $\texttt{Ret.Age}_{t+1}$ depending on the agent's age and current retirement status.

We can see the choice of $\{C_t, \phi_t, \texttt{Ret.Age}_{t+1}\}$ as happening in two steps: the agent first commits himself to a ϕ_t and pays the cost F if necessary and then, conditioning on that choice, picks $\{C_t, \texttt{Ret.Age}_{t+1}\}$. The problem and value function of the agent who takes ϕ_t and his net-of-fixed-cost wealth $\tilde{A}_t$ as given is:

$$\tilde{V}_t(\tilde{A}_t, \phi_t, \texttt{Ret.Age}_t) = \max_{C_t, \texttt{Ret.Age}_{t+1}} u(C_t, \ell) + \beta\delta_t \texttt{Emax}_{t+1}(S_t, \phi_t, \texttt{Ret.Age}_{t+1}) + \not{\delta}\varphi(S_t)$$
$$\text{Subject to:} \quad 0 \leq C_t, \quad 0 \leq S_t, \quad S_t = \tilde{A}_t - C_t.$$

We solve this problem for all combinations of $\tilde{A}_t$ in our wealth grid, ϕ_t on its grid, and the feasible $\texttt{Ret.Age}_t$. The values we consider for C_t are our consumption (proportional) grid times $\tilde{A}_t$. We use the results to construct linear interpolators for $\tilde{V}_t(\cdot, \phi_t, \texttt{Ret.Age}_t)$ for every combination of $\{\phi_t, \texttt{Ret.Age}_t\}$.

Finally, we can express the value function as:

$$V_t(A_t, \texttt{Ret.Age}_t) = \max_{\phi_t} \tilde{V}_{t+1}(\tilde{A}_t, \phi_t, \texttt{Ret.Age}_{t+1})$$
$$\text{Subject to:} \quad 0 \leq \tilde{A}, \quad 0 \leq \phi_t \leq 1, \quad \tilde{A}_t = A_t - \mathrm{F} \times \mathbf{1}\left[\phi_t > 0\right].$$

For every feasible $\texttt{Ret.Age}_{t+1}$, we solve this problem at every A_t in our wealth grid and use the results to construct linear interpolators for $V_t(\cdot, \texttt{Ret.Age}_t)$. With these interpolators, we move on to period $t-1$ and repeat the process.

F. Revenues From Social Security Reforms

F.1. The Steady State Cost of Social Security

Parameterizing our counterfactual policies in a way that makes them comparable requires us to calculate what would be the per-period per-person cost of a Social Security system in *steady state*. To calculate this cost, we start by grouping all the age-invariant characteristics of our model—EA score, college attendance, childhood socioeconomic status, birth year, pension arrangement, and unobserved heterogeneity draws—in a vector that we call $\vec{H}_i$ for person i. Then, we use $\mathcal{C}(a, \vec{H})$ to denote the expected net revenue that the government will collect from a person with age-invariant characteristics $\vec{H}$ when he is a-years old through the social security system. We consider the net revenue to be FICA taxes paid on income if working or the negative of social security benefits if retired.

We think of a *steady state* as a point in which: a) the number of agents born each period and the distribution of their age-invariant characteristics, which we will denote with N_{ss} and $F_{\vec{H}}$ respectively, are both constant over time; and b) all of the model's components (shock distributions, earning patterns) have stabilized so that $\mathcal{C}(a, \vec{H})$ is not time-varying.

If we denote p_a the probability that a person survives to age a; we know that in any steady state period there will be $p_a \times N_{ss}$ agents of age a alive. Since in our model survival probabilities are not related to characteristics $\vec{H}$, the distribution of characteristics across agents of any age at any steady-state time will be $F_{\vec{H}}$. Therefore,

Table G4: Implied Sharpe ratios and relative Sharpe ratio losses.

			Percentiles					
Variable	Mean	Std. Dev.	25th	50th	75th	90th	95th	99th
Sharpe Ratio	0.207	0.088	0.134	0.211	0.280	0.324	0.341	0.356
RSRL	0.440	0.239	0.238	0.426	0.635	0.792	0.838	0.908

The table reports summary statistics of the Sharpe ratio of the risky asset that is available to our simulated agents. The relative Sharpe ratio loss (RSRL) measures how far an agent's risky investments are from the maximum Sharpe ratio available, which in our case is that of the S&P500.

the total per-period expected net Social Security revenue will be:

$$\sum_{a=21}^{90}\left(p_a N_{ss}\times\int\mathcal{C}(a,h)dF_{\vec{H}}(h)\right)\propto\sum_{a=21}^{90}p_a\times E_{\vec{H}}[\mathcal{C}(a,\vec{H})].$$

We approximate the right hand side of the previous equation using our simulated sample $\mathcal{I}$: $\sum_{a=21}^{90}\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}c_{i,a}$, where $c_{i,a}$ is the net Social Security revenue collected from agent i at age a (zero if the agent is dead). This approximation relies on assuming that our sample can produce a good approximation of the steady state average social security revenue at every age.

G. Sharpe Ratios Implied by Our Estimates

Our specification of risky-asset log-returns in Equation 7 and its estimated parameters in Table 3 generate differences in the expected returns that agents get from investing in the risky asset, as Figure 2 shows. Since, on the contrary, the volatility of risky asset log-returns does not vary across individuals, there are differences in the expected compensation per unit of risk that they get. This section examines those differences in greater detail.

We use two measures of the risk-adjusted compensation that agents in our model are receiving in their investments: the Sharpe ratio, $\frac{\mathbb{E}[\tilde{R}_{i,t}]-R}{\sqrt{\mathbb{V}(\tilde{R}_{i,t})}}$ and the "relative Sharpe ratio loss" or RSRL (Calvet, Campbell, and Sodini, 2007), $1-\frac{\text{Sharpe}_i}{(\mathbb{E}[R_t^{\text{SP500}}]-R)/\sqrt{\mathbb{V}(R_t^{\text{SP500}})}}$. The Sharpe ratio measures how much of a premium in expected returns agents are getting over the risk-free rate for each unit of return volatility that they take on. The relative RSRL measures how much lower than a benchmark (the S&P 500 index in our case) an agent's Sharpe ratio is in relative terms. Table G4 presents summary statistics of the distribution of Sharpe ratios and RSRLs for the full population of simulated agents. The table shows that agents in our model receive very different compensations for their financial risk-taking and that most of them are far from the performance of the benchmark return R_t^{SP500}. The median agent's available risky investments deliver a compensation per unit of risk that is 42.6% lower than that of the benchmark S&P500, and one tenth of agents earn compensations almost 80% lower. We cannot directly construct observed Sharpe ratios and RSRLs for the HRS sample since doing so would require more detailed information on their investments. However, the ranges we calculate using the estimated model are similar to those estimated by Calvet, Campbell, and Sodini (2007) using Swedish administrative data.[7]

References

B. W. Arold, P. Hufe and M. Stoeckli. 2022. "Genetic Endowments, Educational Outcomes and the Mediating Influence of School Investments." *CESifo Working Papers* (9841).

S. Baker, P. Biroli, H. van Kippersluis et al. 2022. "Beyond Barker: Infant Mortality at Birth and Ischaemic Heart Disease in Older Age."

S. H. Barcellos, L. Carvalho and P. Turley. 2021. "The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages, and Later-Life SES."

S. H. Barcellos, L. S. Carvalho and P. Turley. 2018. "Education Can Reduce Health Differences Related to Genetic Risk of Obesity." *Proceedings of the National Academy of Sciences* 115 (42):E9765–E9772.

D. Barth, N. W. Papageorge and K. Thom. 2020. "Genetic Endowments and Wealth Inequality." *Journal of Political Economy* 128 (4):1474–1522.

J. P. Beauchamp, D. Cesarini, M. Johannesson et al. 2011. "Molecular Genetics and Economics." *Journal of Economic Perspectives* 25 (4):57–82.

D. W. Belsky, B. W. Domingue, R. Wedow et al. 2018. "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences* 115 (31):E7275–E7284.

D. W. Belsky, T. E. Moffitt, D. L. Corcoran et al. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment

[7] Calvet, Campbell, and Sodini (2007) estimate RSRLs whose 25th and 99th percentiles are {0.29,0.89}, {0.07,0.85}, or {-0.16,0.82} depending on the benchmark used for comparison.

Relate to Life-Course Development." *Psychological Science* 27 (7):957–972.

D. J. Benjamin, D. Cesarini, C. F. Chabris et al. 2012. "The Promises and Pitfalls of Genoeconomics." *Annual Review of Economics* 4 (1):627–662.

P. Biroli. 2015. "Genetic and Economic Interaction in Health Formation: The Case of Obesity." *Working Paper* :70.

L. E. Calvet, J. Y. Campbell and P. Sodini. 2007. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." *Journal of Political Economy* 115 (5):707–747.

J. M. Fletcher and D. Conley. 2013. "The Challenge of Causal Inference in Gene–Environment Interaction Research: Leveraging Research Designs From the Social Sciences." *American Journal of Public Health* 103 (Suppl 1):S42–S45.

J. M. Fletcher and Q. Lu. 2021. "Health policy and genetic endowments: Understanding sources of response to Minimum Legal Drinking Age laws." *Health Economics* 30 (1):194–203.

G. Guo and E. Stearns. 2002. "The Social Influences on the Realization of Genetic Potential for Intellectual Development*." *Social Forces* 80 (3):881–910.

K. P. Harden and P. D. Koellinger. 2020. "Using Genetics for Social Science." *Nature Human Behaviour* 4 (6):567–576.

M. Houmark, V. Ronda and M. Rosholm. 2020. "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills." SSRN Scholarly Paper ID 3708642, Social Science Research Network, Rochester, NY.

S. Imrohoroğlu and S. Kitao. 2012. "Social Security Reforms: Benefit Claiming, Labor Force Participation, and Long-Run Sustainability." *American Economic Journal: Macroeconomics* 4 (3):96–127.

J. J. Lee, R. Wedow, A. Okbay et al. 2018. "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals." *Nature Genetics* 50 (8):1112–1121.

A. Lusardi, P.-C. Michaud and O. S. Mitchell. 2017. "Optimal Financial Knowledge and Wealth Inequality." *Journal of Political Economy* 125 (2):431–477.

A. R. Martin, C. R. Gignoux, R. K. Walters et al. 2017. "Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations." *The American*

Journal of Human Genetics 100 (4):635–649.

D. Muslimova, H. van Kippersluis, C. A. Rietveld et al. 2020. "Dynamic Complementarity in Skill Production: Evidence From Genetic Endowments and Birth Order." SSRN Scholarly Paper 3748468, Social Science Research Network, Rochester, NY.

R. E. Nisbett, J. Aronson, C. Blair et al. 2012. "Intelligence: New Findings and Theoretical Developments." *American Psychologist* 67 (2):130–159.

N. W. Papageorge and K. Thom. 2020. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." *Journal of the European Economic Association* 18 (3):1351–1399.

V. Ronda, E. Agerbo, D. Bleses et al. 2022. "Family Disadvantage, Gender, and the Returns to Genetic Human Capital*." *The Scandinavian Journal of Economics* 124 (2):550–578.

S. Scarr-Salapatek. 1971. "Race, Social Class, and IQ." *Science* 174 (4016):1285–1295.

L. L. Schmitz and D. Conley. 2017. "The Effect of Vietnam-era Conscription and Genetic Potential for Educational Attainment on Schooling Outcomes." *Economics of Education Review* 61:85–97.

R. J. Shiller. 1990. *Market Volatility.* Cambridge, MA, USA: MIT Press.

Social Security Administration. 2019. "Annual Statistical Supplement to the Social Security Bulletin, 2018."

S. Trejo and B. W. Domingue. 2018. "Genetic Nature or Genetic Nurture? Introducing Social Genetic Parameters to Quantify Bias in Polygenic Score Analyses." *Biodemography and Social Biology* 64 (3-4):187–215.

E. M. Tucker-Drob and T. C. Bates. 2016. "Large Cross-National Differences in Gene × Socioeconomic Status Interaction on Intelligence." *Psychological Science* 27 (2):138–149.

E. Turkheimer, A. Haley, M. Waldron et al. 2003. "Socioeconomic Status Modifies Heritability of IQ in Young Children." *Psychological Science* 14 (6):623–628.

P. M. Visscher, N. R. Wray, Q. Zhang et al. 2017. "10 Years of GWAS Discovery: Biology, Function, and Translation." *The American Journal of Human Genetics* 101 (1):5–22.

K. I. Wolpin. 2013. *The Limits of Inference without Theory.* The MIT Press.