HCEO WORKING PAPER SERIES

Genetic Endowments and Wealth Inequality

Daniel Barth
Nicholas W. Papageorge
Kevin Thom

Working Paper

2018-077

10/2018

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Genetic Endowments and Wealth Inequality*

Daniel Barth† Nicholas W. Papageorge‡ Kevin Thom§

Abstract: We show that genetic endowments linked to educational attainment strongly and robustly predict wealth at retirement. The estimated relationship is not fully explained by flexibly controlling for education and labor income. We therefore investigate a host of additional mechanisms that could help to explain the gene-wealth gradient, including inheritances, mortality, savings, risk preferences, portfolio decisions, beliefs about the probabilities of macroeconomic events, and planning horizons. The associations we report provide preliminary evidence that genetic endowments related to human capital accumulation are associated with wealth not only through educational attainment and labor income, but also through a facility with complex financial decision-making. Our study illustrates how economic research seeking to understand sources of inequality can benefit from recent advances in behavioral genetics linking specific observed genetic endowments to economic outcomes.

Keywords: Wealth, Inequality, Portfolio Decisions, Beliefs, Education and Genetics.
JEL Classification: D14, D31, G11, H55, I24, J24.

*First version: February 27, 2017. This version: May 23, 2018. We thank Aysu Okbay for constructing and sharing the polygenic scores used for HRS respondents. We are also grateful for helpful comments from Robert Barbera, Daniel Belsky, Lee Benham, Jess Benhabib, Daniel Benjamin, Daniel Bennett, Alberto Bisin, Christopher Carroll, David Cesarini, Gabriella Conti, Stephanie DeLuca, Manasi Deshpande, Weili Ding, Benjamin Domingue, Steven Durlauf, Jon Faust, Titus Galama, Barton Hamilton, Bruce Hamilton, Joseph Hotz, Steven Lehrer, George-Levi Gayle, Robin Lumsdaine, Shelly Lundberg, Luigi Pistaferri, Robert Pollak, Paul Romer, Simone Schaner, Stephan Siegel, Matthew Shapiro, Dan Silverman, Jonathan Skinner, Rachel Thornton, Robert Topel, Jasmin Wertz, Robert Willis, Jonathan Wright, and Basit Zafar, along with seminar participants at CESR, Clemson, CUNY - Baruch College, Dartmouth, Duke, Johns Hopkins, Michigan, Penn State, Rochester, UCSB, UNC-Chapel Hill, UW-Milwaukee, William and Mary, Bates White, the 2016 HCEO Conference on Genetics and Social Science, the 2nd Annual Empirical Microeconomics Conference at ASU, the 2017 NBER Cohort Studies meetings, the 2017 North American summer meetings of the Econometric Society, 2017 NBER Institute (Aging), and the 2017 PAA Meetings. We acknowledge excellent research assistance from Andrew Gray. Finally, the authors would like to note that a previous version of this paper was circulated with the title "Genetic Ability, Wealth and Financial Decision-Making." The usual caveats apply.

†Center for Economic and Social Research, Univ. of Southern California. Email: dannybarth@gmail.com.
‡Johns Hopkins University, IZA and NBER. Email: papageorge@jhu.edu.
§New York University. Email: kevin.thom@nyu.edu.

1 Introduction

Wealth inequality in the United States and many other countries is large and growing (Saez and Zucman, 2014; Jones, 2015). An extensive literature attempts to explain wealth inequality through variation in the earnings process, entrepreneurial talent, bequests, risk aversion, and time-discounting. While these factors can generate a skewed wealth distribution, they fail to reproduce other key empirical features including the thickness of the tail.[1]

A new wave of theoretical work argues that cross-sectional heterogeneity in the returns to wealth is required to match the basic features of the wealth distribution (Benhabib, Bisin, and Zhu, 2011; Benhabib and Bisin, 2016). This argument is supported by a growing empirical literature that finds substantial heterogeneity in such returns (Fagereng et al., 2016; Benhabib, Bisin, and Luo, 2015; Bach, Thiemann, and Zucco, 2015). Much of this heterogeneity persists over time, with some individuals earning consistently higher returns to wealth (Fagereng et al., 2016).[2] Despite its theoretical and empirical importance, little is known about what drives such persistence. However, policy responses to wealth inequality are likely to have different effects depending on the mechanisms through which persistent heterogeneity operates — for example, whether this heterogeneity comes from variation in preferences (e.g. risk aversion) or variation in skills (e.g. financial decision-making).

In this paper, we identify a biological basis for heterogeneity in wealth: genetic endowments related to human capital accumulation. In particular, we make use of molecular genetic data in the Health and Retirement Study (HRS) along with recently estimated associations between genetic markers and educational attainment. We show that the same observed genetic markers that predict education also predict household wealth in the HRS. Importantly, the estimated association between these markers and wealth is economically large and statistically significant after controlling for education and labor market income, suggesting that these variants operate through additional channels. Indeed, we show that these markers predict parental transfers such as inheritances, mortality, and risk preferences. We also find they are related to dimensions of financial decision-making, including stock market participation, business ownership, the financial planning horizon, and macroeconomic expectations.

While we do not observe returns directly, our results provide a possible genetic microfoundation for the persistent differences in returns to wealth needed to explain existing wealth

[1]Some models can match the thickness of the tail, but only under implausible assumptions about the level of heterogeneity in the earnings process. For example, Kindermann and Krueger (2014) require the top 0.25% of earners to earn 400-600 times more than the median earner. Empirically, this number is closer to 33.

[2]In an earlier contribution, Yitzhaki (1987) finds that individuals with higher labor market income invest in better performing assets. Because higher income is associated with higher wealth, this offers indirect evidence of a positive correlation between wealth and the returns to wealth.

distributions. The genetic transmission of traits related to returns to wealth may thus provide another source for the well-documented intergenerational persistence of wealth (Charles and Hurst, 2003; De Nardi, 2004; Benhabib, Bisin, and Zhu, 2011). An understanding of the intergenerational transmission of economic outcomes that does not account for the role of genetics is likely to be incomplete, placing too much weight on other factors such as parental investments and financial transfers.[3] Moreover, because financial decisions and beliefs about the macroeconomy play a role in explaining this gradient, relatively straightforward policy tools such as stronger public pension schemes may help to reduce wealth inequality stemming from genetic variation. This is especially relevant given the dramatic shift away from defined-benefit retirement plans towards options that give individuals greater financial autonomy (Poterba and Wise, 1998).

Our work contributes to an existing literature on endowments, economic traits, and household wealth. One strand of this work examines how various measures of "ability," such as IQ or cognitive test scores, predict household wealth and similar outcomes (Grinblatt, Keloharju, and Linnainmaa, 2011; Grinblatt et al., 2015; Lillard and Willis, 2001).[4] However, parental investments and other environmental factors can directly affect test performance, making it difficult to separate the effects of endowed traits from endogenous human capital investments. A second strand of this literature focuses on genetic endowments, and seeks to estimate their collective importance using twin studies. Twin studies have shown that genetics play a non-trivial role in explaining financial behavior such as savings and portfolio choices (Cronqvist and Siegel, 2014, 2015; Cesarini et al., 2010).[5] However, while twin studies can decompose the variance of an outcome into genetic and non-genetic contributions, they do not identify which particular markers influence economic outcomes.[6] Moreover, it is typically impossible to apply twin methods to large and nationally representative longitudinal studies, such as the HRS, which offer some of the richest data on household wealth and related

[3]On the importance of inter-generational transmission of economic outcomes, a key contribution is Black, Devereux, and Salvanes (2005).

[4]As we discuss in greater detail in Section 2, when describing the genetic endowments examined in this paper we purposefully avoid the term "ability" because it is likely overly simplistic and imprecise. For example, the term does not emphasize multidimensionality of skill. The genetic endowments we study, which predict educational attainment, may capture some types of cognitive skill, but may also capture a host of other factors, such as personality or socio-emotional skills.

[5]For example, using the Swedish Twin Registry, Cesarini et al. (2010) demonstrate that about 25% of individual variation in portfolio risk is attributable to genetic variation while Cronqvist and Siegel (2015) show that 35% of variation in propensity to save has a genetic basis. It is worth mentioning, however, that these estimates may be biased upward if identical twins face more similar family environments than do non-identical twins (Fagereng et al., 2015).

[6]Variance decomposition exercises such as twins studies treat genes as unobserved factors. Testing hypotheses about specific mechanisms is conceptually possible using information on twins. In practice, learning about interactions between observed and unobserved factors is generally difficult and relies on modeling assumptions and requires large amounts of data to permit stratification by each potential mediating factor.

behavioral traits.

The development of *polygenic scores* — predictive indices that aggregate many genetic markers — offers a new approach to studying endowments that overcomes some of the limitations of test scores and twin methods. Results from the state of the art in this literature (Okbay et al., 2016; Lee and et al, 2018) allow for the construction of a polygenic score that robustly predicts education, including in the HRS (Papageorge and Thom, 2016). Unlike traditional proxies such as cognitive test sores (e.g. IQ), genetic measures are *predetermined* if not exogenous, and are not causally affected by endogenous investments. Moreover, directly observing these traits as an observable index in the HRS greatly eases the process of assessing the mechanisms through which these endowments operate and estimating how they interact with policies and macroeconomic conditions.

We first establish a robust relationship between household wealth in retirement and the average household polygenic score for educational attainment.[7] A one-standard-deviation increase in the score is associated with a 33.1 percent increase in household wealth (approximately $144,000 in 2010 dollars). The relationship between the polygenic score and wealth is present across time and education groups. Measures of educational attainment, including years of education and completed degrees, explain over half of this relationship. Using detailed income data from the Social Security Administration (SSA) as well as self-reported labor earnings from the HRS, we find that labor income can explain only a small part of the gene-wealth gradient that remains after controlling for education.[8] These results indicate that while education and labor market earnings are important sources of variation in household wealth, they explain only a portion of the relationship between genetic endowments and wealth.[9]

Next, we explore additional mechanisms that may explain the gene-wealth gradient. We first examine intergenerational transfers and motivations for higher savings. We show that the polygenic score is positively related to the probability of receiving an inheritance, but not the size of the inheritance conditional on receiving one. We also find that individual polygenic scores predict the objective probability of death as well as subjective beliefs about mortality. Because higher expected longevity may lead to higher savings rates, we then test

[7]In Section 3, we explain why we use a average household score and also discuss robustness to alternative specifications.

[8]As we explain in greater detail below when describing the measure of genetic endowments used in this paper and when presenting results, all variables we use are measured with error, which means care must be taken in how we interpret estimates. While controlling for income and earnings variables available for HRS respondents does not fully explain the gene-wealth gradient, it is possible that less noisy measures of both would fully explain the gradient.

[9]Though higher lifetime-income households save more (Dynan, Skinner, and Zeldes, 2004), as Venti and Wise (1998) demonstrate using HRS data, lifetime income does not fully explain wealth inequality.

whether higher polygenic score households save more, but we find no empirical evidence of such a mechanism.[10]

While we find no relationship between the polygenic score and savings rates, we do find a relationship with *how* households invest and save. We first document an association between the polygenic score and measures of risk tolerance constructed from responses to hypothetical income and wealth gambles.[11] Higher polygenic score households are also more likely to own businesses and invest in the stock market. Motivated by the findings on stock market participation, we next analyze aspects of financial decision-making. We show that lower polygenic scores predict the reporting of beliefs about the probabilities of macroeconomic events that are at odds with objectively correct probabilities and that are heaped on probabilities of 0 percent or 100 percent (a phenomenon we refer to as "extreme beliefs"). Large deviations between subjective and objective probabilities may reflect difficulty with probabilistic thinking. We also find that higher polygenic score households report longer planning horizons for financial decisions. This may indicate that these households are more patient, or that they are more comfortable with complex and abstract decision problems and therefore adopt longer planning horizons.

We end our study by examining how the polygenic score interacts with a policy relevant variable: pensions. Because defined-benefit pensions offer recipients a guaranteed stream of income without requiring them to make choices about contribution rates or asset composition, such plans should reduce differences in wealth that arise from skill in financial decision-making. We find that the gene-wealth gradient is over five times as large for the subset of households who do not participate in defined-benefit pension plans. This exercise is useful for two reasons. First, it offers compelling support for the hypothesis that financial decisions may be a source of the gene-wealth gradient. Second, it also highlights a potentially important policy consideration. While more flexible plans like 401(k) accounts grant individuals greater freedom in planning for retirement, they may also reduce the welfare of those who find it more difficult to navigate complex financial choices.

The remainder of this paper is organized as follows. Section 2 provides details on the genetic index used in this paper, as well as its use in economics. Section 3 introduces the data and provides details on key variables used in this study. Section 4 presents our main results on the relationship between the average household polygenic score and household wealth. Section 5 explores a host of possible mechanisms that can explain the gene-wealth

[10]This is related to the findings of Cronqvist and Siegel (2015), who use a twins design to study a genetic basis for savings behavior. However, they find that genes related to savings do not operate through genes related to education, but instead through time preference and self control because of genetic correlations between savings, smoking, and obesity.

[11]Previous research suggests that risk aversion has a genetic basis (Cesarini et al., 2009).

gradient, including standard factors established in the literature along with measures of financial decision-making. Section 6 concludes.

2 Molecular Genetic Data and Economic Analysis

Following recent developments in the genetics literature, we summarize the genetic factors related to educational attainment using a linear index known as a polygenic score. In this section, we first provide details on the construction of the particular polygenic score that we use. We then discuss what this approach can add to economic analysis.

2.1 Constructing the Polygenic Score

2.1.1 The Human Genome

The human genome consists of approximately 3 billion nucleotide pairs spread out over 23 chromosomes.[12] A DNA molecule is often thought of as double-helix ladder, with the nucleotide base pairs forming the "rungs" of the ladder. Each rung can either be an adenine-thymine (AT) pair, or a guanine-cytosine (GC) pair. Thus, the human genome can be thought of as a series of 3 billion genetic addresses, each of which contains a particular base pair molecule in a particular position.[13] Furthermore, each individual possesses two copies of each chromosome, one from each parent. A gene refers to a particular subsequence of base pairs within a chromosome which contain instructions for a particular enzyme or functional molecule. Genes vary in length and may consist of thousands of base pairs.

At the vast majority of locations in the genome (about 99%), there is no variation across individuals in the nucleotide pair that is present. At the remaining locations (less than 1%), the base pair may differ across individuals.[14] Such locations are referred to as single-nucleotide polymorphisms (SNPs, pronounced "snips"). Particular SNPs are referred to by

[12]Much of the background information presented here on the human genome follows Beauchamp et al. (2011) and Benjamin et al. (2012).

[13]If the DNA strand can be thought of as a ladder with the nucleotide pairs being the rungs, then the rails or sides of the ladder are formed by phosphate and sugar molecules. These rails can be distinguished as either the positive (+) or negative (−) strands. At a particular location, it will matter which nucleotide molecule is attached to which strand. For example, if there is an adenine-thymine pair in a particular position where the adenine molecule is attached to the positive strand, this would be denoted by an A. However, if instead the thymine molecule were attached to the positive strand, this would be denoted by a T. This means that four possible variants could exist at a given address: A, T, G, or C, depending on which nucleotide pair is present, and the position of that pair relative to the positive strand. However, most SNPs are *biallelic*, meaning that there are only two observed alleles at a particular location.

[14]We also note that other forms of genetic variation exist. Such variation is typically referred to as structural variation and may include deletions, insertions, and copy-number variations. (Feuk, Carson, and Scherer, 2006).

a name (e.g. rs7937), which indicates its position in the genome. For each individual and at each SNP location, we can typically record an individual's genotype by counting the number of copies of a reference molecule (also called an allele, either AT or GC) that an individual possesses at a particular location. Since each individual has two copies of each chromosome, an individual can either have zero, one, or two instances of the reference allele. Molecular genetic data thus most commonly take the form of a series of count variables indicating the number of copies of the reference allele possessed by an individual at a particular SNP (e.g. $rs7937_i \in \{0, 1, 2\}$). The maximum value of 2 reflects the fact that an individual can have at most two copies of the reference allele — one from each parent. A major task of behavioral genetics involves determining which, if any, of these SNP variables are associated with behavioral outcomes.

2.1.2 Genetic Studies, GWAS, and Polygenic Scores

Twins studies account for much of the existing literature on genetics and economic behaviors. A standard twins methodology estimates the fraction of the variance of a particular outcome due to genetic factors by comparing the outcomes of monozygotic (identical) twins and dizygotic (fraternal) twins. While monozygotic twins share nearly all SNPs in common, dizygotic twins will share only about 50 percent of SNPs. Twins studies often assume the following data generating process for an outcome of interest, Y_{if} for individual i in family f:

$$Y_{if} = A_i + C_f + E_i \tag{1}$$

Here A_i represents an *additive* genetic component, C_f represents *common* environmental factors affecting all individuals in family f, and E_i represents unique *environmental* factors affecting individual i. Differences in the covariance of Y_{if} between monozygotic and dizygotic twins allows one to identify the heritability of this outcome, which is the fraction of the variance of Y_{if} accounted for by genetic differences: $\frac{Var(A_i)}{Var(Y_{if})}$. Existing twins studies deliver heritability estimates of around 40% for education (Branigan, McCallum, and Freese, 2013).

While twins studies provide an estimate of how much genetic factors collectively matter for explaining variation in a given trait, they do not reveal which specific SNPs are relevant. By contrast, Genome Wide Association Studies (GWAS) scan the entire genome and estimate associations between individual SNPs and outcomes of interest. A GWAS typically proceeds by gathering genotypic data on J observable SNPs $\{SNP_{ij}\}_{j=1}^{J}$ and estimating J separate regressions similar to the following:

$$Y_i = \mu X_i' + \beta_j SNP_{ij} + \epsilon_{ij} \tag{2}$$

Here SNP_{ij} measures the number of copies of the reference allele possessed by individual i for SNP j, and X_i is a vector of control variables.[15] A standard practice is to include several principal components of the full matrix of SNP data $\{SNP_{ij}\}_{j=1}^{J}$ as essential controls. One concern is that associations between a SNP and an outcome could reflect population stratification. For example, if a particular SNP is more common in a specific ancestry group (e.g. Southern Europeans), then an observed association between this SNP and the outcome of interest might reflect a combination of the biological function of the SNP and the common environment shared by this ancestry group. Variation in the principal components captures variation in genetic ancestry groups, so controlling for these factors helps account for population stratification (Price et al., 2006). In this sense, the principal components help to control for ethnic background factors that would be absorbed by family fixed effects in research designs exploiting within-family variation. Unless otherwise noted, the principal components are always included in our regression specifications.

The J individual regressions in a GWAS produce a set of coefficients, $\left\{\widehat{\beta}_j\right\}_{j=1}^{J}$, with associated standard errors and p-values. Researchers interested in studying individual genetic markers typically focus on those SNPs exhibiting the strongest GWAS associations.[16] Since traits like education are influenced by a large number of genetic markers with small influences, GWAS results are often used to construct *polygenic scores* — predictive indices formed by linear combinations of SNPs. A polygenic score for a trait or outcome of interest is therefore given by:

$$PGS_i = \sum_{j=1}^{J} \widetilde{\beta}_j SNP_i \tag{3}$$

Here, the weights $\left\{\widetilde{\beta}_j\right\}_{j=1}^{J}$ represent transformations of the original GWAS coefficients, $\left\{\widehat{\beta}_j\right\}_{j=1}^{J}$. This predictive index can be thought of as the best SNP-based linear predictor of the common genetic component A_i from the linear model specified in Equation (1). Using the unadjusted coefficients as the PGS weights could reduce the accuracy of the index in the presence of correlation between SNPs (called linkage disequilibrium, or LD in the genetics literature), or if the coefficients are imprecisely estimated. One method for adjusting these

[15] Separate regressions for each SNP are estimated because in practice, one typically has many more genotyped SNPs than observations in a discovery sample.

[16] Given the large number of regression equations being estimated, correction for multiple hypothesis testing has been a key concern in this literature. For the purposes of determining whether an individual SNP-outcome association is statistically significant, the literature has adopted stringent p-value thresholds. A benchmark threshold for *genome wide significance* is $p < 5 \times 10^{-8}$. Stringent thresholds were developed in part as a response to earlier methods used to measure gene-outcome associations using so-called *candidate genes*, which are genes that the researcher believes may be implicated in an outcome arising from knowledge of biological processes. This approach suffered from false positives due to an uncorrected multiple-hypothesis testing problem (Benjamin et al., 2012).

coefficients and constructing PGS weights — the one used to construct the score that we study — is the LDpred procedure of Vilhjalmsson (2015). LDpred weights are constructed through a Bayesian approach that updates a prior given the GWAS results $\left\{\widehat{\beta}_j\right\}_{j=1}^{J}$ and the covariance matrix of the SNPs (the LD matrix).[17] The LDpred weight $\widetilde{\beta}_j$ is the expected value of the resulting posterior distribution.

GWAS results, and gene discovery studies more generally, have largely focused on medical or health-related outcomes including smoking (Bierut, 2010; Thorgeirsson et al., 2010) and obesity (Locke et al., 2015). However, the increasing availability of molecular genetic data has made it possible to perform well-powered GWAS for behavioral traits with more distant relationships to underlying biological mechanisms. In particular, a series of landmark studies have delivered the first GWAS associations between individual SNPs and educational attainment (Rietveld et al. (2013); Okbay et al. (2016); Lee and et al (2018)). Existing work shows that polygenic scores for educational attainment based on these GWAS predict labor market outcomes including earnings (Papageorge and Thom, 2016), and other measures of adult success (Belsky et al., 2016), even after controlling for completed education.

In this paper, we study a polygenic score based on the educational attainment GWAS results from Lee and et al (2018), which featured a discovery sample of over 1.1 million people. The score is constructed with the LDpred method, using parameters outlined in Okbay, Benjamin, and Visscher (2018).[18] We use a score based on the LDpred method because it has been shown to increase out of sample predictive power relative to alternate approaches such as pruning and trimming (Okbay et al., 2016).[19] Importantly, HRS data

[17]The prior used for LDpred puts probability p on the event that any SNP has a non-zero association with the outcome, and conditional on having a non-zero association, it is assumed that the true SNP coefficients are distributed $\beta_j \sim N\left(0, \frac{h_g^2}{Mp}\right)$, where the variance depends on the heritability of the trait, h_g^2, the number of SNPs used in the score, M, and the assumed probability that a SNP has a non-zero association, p.

[18]Specifically, the score is based on GWAS associations for 1,104,681 SNPs that pass the inclusion criteria documented in Okbay, Benjamin, and Visscher (2018). The LD matrix is estimated using 8,353 genetically European individuals in the HRS passing certain quality control measures for the genetic data. When correcting a SNP association for possible correlation with other SNPs, one considers possible correlations between the SNP and other SNPs within a particular neighborhood of that SNPs. The size of this neighborhood (number of adjacent SNPs considered in either direction of the SNP) is given by the LD window. The score we use is constructed with an LD window of $M/3000$, where M is the number of included SNPs. The prior used to construct the score assumes that there is a probability $p = 1$ that a SNP is has a non-zero association. It is assumed that non-zero associations are distributed normally with mean zero, and a variance equal to $\frac{h_g^2}{Mp}$, where h_g^2 is a heritability estimate derived from the GWAS summary statistics for each SNP (Vilhjalmsson, 2015), p is the assumed probability of a non-zero association, and M is the number of SNPs included in the analysis.

[19]We note that there are several alternative methods for constructing the weights used for polygenic scores. One common strategy for adjusting these coefficients, called *pruning and trimming*, sets $\widetilde{\beta}_j = 0$ for SNPs that are highly correlated with other SNPs, as well as those with high p-values. We discuss robustness to alternative methods to construct scores in the Appendix.

are not used to estimate the GWAS associations $\left\{\widehat{\beta}_j\right\}_{j=1}^{J}$ for this score, so every analysis in this study is an out-of-sample exercise. Prediction results from Lee and et al (2018) suggest that this score explains approximately 10.6 percent of the variation of years of schooling in the Health and Retirement Study. In what follows, we refer to this score as the Educational Attainment or *EA score*.[20]

2.1.3 Interpreting Variation in the EA Score

In the absence of measurement error or gene-environment correlations, we could interpret variation in the EA score as reflecting differences across individuals in genetic endowments that influence personal traits that promote educational attainment. Due to measurement error, one cannot use the EA score to draw conclusions about the total amount of variation in an outcome that is attributable to genetic endowments. This is one drawback of the use of polygenic scores relative to twin studies.

It is reasonable to suspect that genetic endowments related to educational attainment may affect biological processes related to cognition that facilitate learning. Indeed, pathway analyses suggest that several of the SNPs most heavily tied to educational attainment are linked to biological processes known to be involved in brain development and cognitive processes (Lee and et al, 2018; Okbay et al., 2016). Moreover, results from Belsky et al. (2016) suggest that an earlier polygenic score for educational attainment predicts cognitive test scores for children in elementary school. However, it is important to note that the GWAS associations can reflect a range of traits — both cognitive and non-cognitive — that affect educational attainment through diverse mechanisms. The EA score may very well reflect personality traits and other attributes in addition to cognition.

One of the largest challenges in interpreting variation in the EA score comes from gene-environment correlations. In particular, because individuals inherit their genetic material from their parents, and those parents shape childhood environments, differences in the EA score could reflect not only differences in biological traits that promote educational attainment but also environmental factors that affect education and other outcomes regardless of one's biology. Lee and et al (2018) find that associations between SNPs and educational attainment tend to be smaller using only within-family variation as opposed to within and across family variation. The SNPs that enter the EA score may therefore operate partially through environmental channels such as parental resources and sibling spill-overs. While some of our empirical specifications include controls for parental education, this is likely an

[20] We maintain this nomenclature to distinguish this polygenic score from others that have been constructed to summarize genetic endowments related to different outcomes, such as depression, smoking or subjective well-being.

incomplete measure of investments or rearing environments. Thus, an important limitation of our analyses is that we are not able to cleanly separate the association between the EA score and wealth into biological and environmental components.

Finally, we caution against interpreting variation in the EA score across ethnic ancestry groups. When computing the polygenic score in the HRS sample, we only consider individuals of genetic European ancestry as categorized by the HRS.[21] The EA score we study was constructed using results from a sample of individuals of European ancestry.[22] Previous work has shown that polygenic scores based on GWAS of genetic Europeans lack predictive power and in some cases can generate bizarre predictions when applied to non-European sub-samples. For example, applying a polygenic score for height discovered on a sample of individuals of European descent predicts very low average height relative to the observed distribution if applied to individuals of African descent (Martin et al., 2017).[23] It would thus be misleading and irresponsible to use a polygenic score for education to make predictions about individuals who are not of European descent.

2.1.4 Polygenic Scores and Economic Analysis

Polygenic scores like the EA score described here have many potentially useful applications in economic analysis. As Beauchamp et al. (2011) note, such scores could be used to control for traditionally unobserved characteristics when estimating the effects of an intervention. These could improve the precision of an estimated effect, or reduce bias in the case of non-experimental studies. The scores could also be useful in uncovering treatment effect heterogeneity on the basis of biological factors. Schmitz and Conley (2017) offer a useful example, demonstrating that the effect of Vietnam conscription on educational attainment varied on the basis of an earlier EA score.

Our approach is to view the polygenic score as providing information about the structure of heterogeneity in the economy, and therefore as an important economic object in its own right. Our goal in studying the EA score and its relationship with wealth is to better understand whether previously unobserved biological factors influence household wealth accumulation, and to learn about which economic mechanisms (e.g. income, risk-aversion, decision-making, etc.) may link these factors with household outcomes.

[21]As part of the genetic data release, the HRS also released a file flagging 8,652 individuals as being of European descent based on their genetic markers. The HRS defines such individuals as "... all self-reported non-Hispanic whites that had [principal component] loadings within ± one standard deviations of the mean for eigenvectors 1 and 2 in the [principal components] analysis of all unrelated study subjects."

[22]It is worth mentioning that our results are virtually unchanged if we instead restrict our attention to any HRS individuals who self-identify as being of European descent.

[23]The authors write, "the African populations sampled are genetically predicted to be considerably shorter than all Europeans and minimally taller than East Asians, which contradicts empirical observations (p. 7)"

Our use of the EA score as a measure of biological traits linked to human capital is related to previous attempts in the literature to measure *ability* through the use of tests scores such as IQ or the AFQT. Ability is often interpreted as the innate (pre-investment) characteristics of an individual that influence human capital accumulation and the returns to human capital investments (Becker, 1964). We note two important differences between the EA score and a measure like IQ that make it valuable to study polygenic scores. First, a polygenic score like the EA score can overcome some interpretational challenges related to IQ and other cognitive test scores. Environmental factors have been found to influence intelligence test results and to moderate genetic influences on IQ (Tucker-Drob and Bates, 2015). It is true that differences in the EA score may reflect differences in environments or investments because parents with high EA scores may also be more likely to invest in their children. However, the EA score is fixed at conception, which means that post-birth investments cannot causally change the value of the score. A measure like IQ suffers from both of these interpretational challenges. High IQ parents might have high IQ children because of the genes that they pass on, but also because of the positive investments that they make. That is, since IQ is not determined at conception, it reflects a) individual biological traits, b) correlations between these biological traits and environments, and c) the direct influence of parental investments on IQ test performance. This third channel represents a source of endogeneity bias that is not present in the use of the EA score.

Compared to a cognitive test score like IQ, the EA score may also measure a wider variety of relevant endowments. This is especially important given research, including relatively recent papers in economics, emphasizing the importance of both cognitive and non-cognitive skills in shaping life-cycle outcomes (Heckman and Rubinstein, 2001). Existing evidence suggests a correlation of approximately 0.20 between a cognitive test score available for HRS respondents and the EA score (Papageorge and Thom, 2016). This relatively modest correlation could arise if both variables measure the same underlying cognitive traits with error, or if they measure different traits. However, Papageorge and Thom (2016) find that the relationship between the EA score and income differs substantially from the relationship between later-life cognition scores and income, suggesting that the EA score contains unique information.[24]

In light of the issues discussed in the section, we interpret the EA score as measuring a basket of genetic factors that influence traits relevant for human capital accumulation. They could relate to the type of cognitive function captured by test scores, but may also

[24]The cognitive test administered to HRS respondents was designed to measure cognitive decline among aging adults, which underscores the point that it measures different factors than the EA score. We discuss this point more in Section 5 when we discuss mechanisms explaining the gene-wealth gradient.

include socio-emotional skills and other personality traits. Learning more about how these endowments are linked to economic behavior helps us to understand more about the genetic factors that are important for education. Moreover, understanding how these endowments operate, including their associations with other economic behavior and outcomes, sheds light on the structure of heterogeneity in the economy. The fact that these endowments are genetic matters because their transmission will account for some of the intergenerational persistence of inequality in outcomes, including wealth.

3 The HRS Sample and Key Economic Variables

In this section we introduce data on the key economic variables we study, including demographics and family structure, wealth, and income, among others. We also discuss our unit of analysis (the household) and provide summary statistics for relevant variables. Descriptive statistics for the genetic data are provided in Section 4.

3.1 Sample Construction

The Health and Retirement Study (HRS) is a longitudinal panel study that follows Americans over age 50 and their partners. Surveys began in 1992 and occur every two years. The HRS collected genetic samples from just under 20,000 individuals over the course of four waves (2006, 2008, 2010, 2012). Our sample includes only those genotyped in the 2006 and 2008 waves, since the 2012 wave is not currently available, and the polygenic score we use has not yet been constructed for the 2010 wave.[25] Individuals in the genotyped sample tend to be born in younger birth cohorts because survival until at least 2006 is required for inclusion. Moreover, women and individuals with more education were more likely to agree to the collection of genetic data.[26]

We construct our main analysis sample by first considering all households in which we observe at least one individual who is classified as a genetic European by the Health and Retirement Study. We drop households in which any member self-identifies as non-white.[27] We further restrict our sample to include only retired households in years 1996, 1998, and

[25]A less powerful score based on the results of Okbay et al. (2016) is available for respondents in the 2006-2010 waves. In Appendix B, we show that the main association between this earlier score and wealth is robust to the inclusion of the 2010 sample (see Table S11).

[26]In Appendix A we provide summary statistics on differences between genotyped versus non-genotyped HRS respondents.

[27]We also exclude from the sample male-male and female-female households, as well as households with more than two members, because these households comprise very few observations and may indicate unique circumstances that are not adequately captured by our empirical analysis.

2002-2010.[28] This restriction aims to balance concerns about measurement error in wealth with concerns about selection biases that arise if we drop too many observations from the analysis. Further details on wealth data, including measurement problems, are found in the following section.[29] The resulting analytic sample includes 4,297 households and 15,670 household-year observations, with responses supplied for an average of 3.6 waves.[30]

3.2 Unit of Analysis: Retired Households

Wealth in the HRS is reported at the household level, as are other economic variables of interest such as home ownership and stock market participation. Therefore, our main unit of analysis is the household. Of the 4,297 households we examine, there are three types: 3,026 households that are "coupled," which means both a male and a female are present in at least one wave of the sample, 306 households with only a female member present for all sample observations of that household, and 957 households with only a male member present for all sample observations of that household.

However, our measure of genetic endowments is calculated for individuals rather than households. How best to aggregate EA scores within two-person households is not straightforward. In the empirical analysis, our approach is to take the simple average of EA scores within the household (that is, the average of the two household member scores), but to include individual-level covariates (separately for male and female household members) as controls in the regressions. This is only one of many possible approaches, such as including separate EA scores (for each gender or for both the financial respondent and the non-financial respondent), or using the maximum or minimum score within the household. In Appendix B, we repeat our main specification using each of these alternative approaches and show that our results are qualitatively unchanged across specifications. For the remainder of our

[28] A household is categorized as "retired" if every member of the household is either not working for pay or reports that they are retired. This raises the possibility that some households are included in the sample because they are unemployed, even if they are not retired. This is unlikely to affect our sample given the age of the HRS respondents.

[29] In a series of robustness checks, discussed in greater detail below, we assess the robustness of our main results to alternative definitions of wealth and to alternative sample construction restrictions. In general, we find that our main results are robust to these alternatives. Results from robustness checks are available in Appendix B.

[30] We use household to refer to sets of individuals who cohabitate and report common household-level wealth information. We note, however, that households in the HRS could be connected if, for example, one household splits as the result of divorce to become two separate households with distinct finances. We refer to a family as a set of households that are linked in this way. For example in the event of a divorce, a once-married couple would be observed as one household before the divorce, but would be observed as two separate households after the divorce. Together we consider these households as one family, though they would contribute three separate households to our data set. In subsequent analyses we cluster standard errors at the family level. Of the 4,297 households in our analytical sample, there are 4,182 distinct families.

analysis, unless otherwise noted, the EA score refers to the average household EA score.

Family structure could affect our outcomes of interest, either mechanically through household composition or through selection on unobservables. As a robustness check, we re-estimate our main wealth results including only households that had two members at some point in the sample. Our main results are unchanged by including only two-member households, and we therefore choose to use the full sample of households in our primary specifications. A related issue arises when only one member of the household is genotyped. In such cases, we simply use that household member's genetic information. In Appendix B, we show that these households can be excluded without changing results.

Finally, our main specifications include only retired households. Defined benefit pension income, which can be an important source of household resources at retirement, is poorly measured for individuals who are still working and not drawing pension income. However, we show in a robustness test that our results are robust to including both retired and non-retired households, and we discuss the details of our wealth measures below.

3.3 Description of Key Economic Variables

3.3.1 Summary Statistics for Basic Controls

Table 1 offers summary statistics for four groups of households: (1) all households, regardless of household structure, (2) coupled households, i.e., those with both a male and a female present in at least one wave of the sample, (3) households with only a female member present for all sample observations of that household, and (4) households with only a male member present for all sample observations of that household.

The first two rows of Table 1 show male and female birth years, respectively. Average birth years in the sample range between 1931 and 1935 across the different household structures, with standard deviations between 8.7 and 10.7 years. The next two rows show the mean and standard deviations of years of schooling for males and females, respectively. Across households structures, average years of schooling are between 12.5-12.8 years. The remaining rows report the fraction of males and females that comprise each degree category: no degree, GED, high school, two-year, four-year, masters, professional, unknown/some college. Across household structures, roughly 21% of females and 27% of males attain at least some post-secondary education.

3.3.2 Household Income

Our primary source of earned income data comes from the *Respondent Cross-Year Summary Earnings* data set in the HRS. This file links individuals in the HRS with income data available through the Master Earnings File (MEF) maintained by the Social Security Administration (SSA). The MEF is constructed using data from employers' reports as well as Internal Revenue Service records including W-2 forms and other annual tax figures. The data include "regular wages and salaries, tips, self-employment income, and deferred compensation" (Olsen and Hudson, 2009).[31] The *Respondent Cross-Year Summary Earnings* provides annual MEF income totals for individuals over the period 1951-2013. Our baseline income measure is the sum of all earned income in the MEF associated with a household for all available years through 2010. This may include earnings from deceased spouses that are not directly observed in the HRS.[32]

The SSA data contain earnings information for the entirety of respondents' working lives. This offers a clear advantage relative to the self-reported income measures in the HRS, which only cover older ages. However, an important limitation of the SSA data is that they are top-coded at the taxable maximum amount for Social Security payroll taxes. This taxable maximum has changed substantially over time. In some years, especially in the 1960s and 1970s, a substantial portion of households fall into this category since the maximum was fairly low. For example, in 1965 the maximum was $4,800 (which is about $38,000 in 2018 dollars). In our sample, the average number of person-years during which one income observation was top-coded is 12. Less than one-third (29%) of households are never top-coded. To partially correct for top-coding, we use Current Population Survey (CPS) data to calculate mean income for people earning at least the top-coded level in each year over the period 1961-2010. We then replace a top coded amount in the SSA data with the conditional mean from the CPS data for each of these years.[33] This is admittedly an imperfect solution. As we discuss below, our main results are robust to different ways of treating top-coded income observations.

[31]Olsen and Hudson (2009) offer a detailed discussion of the evolution of the MEF, including the variety of records used to construct annual income in the file, as well as an account of how the kinds of income included in the MEF changed over time.

[32]For each year, we add observed earnings for an individual with any earnings reported for a deceased spouse in the *Deceased Spouse Cross-Year Summary Earnings* data set. After converting annual totals to real 2010 dollars, we then sum up all person-year income observations for each person in a household up through 2010.

[33]For example, if an individual earned $10,000 (nominal) in 1965, we would observe a top-coded income amount of $4,800 in the SSA file. The mean CPS income for those earning at least $4,800 in 1965 is $8,103 so we would replace this individuals' income (any 1965 SSA amount of at least $4,800) with $8,103, which is approximately $56,096 in 2010 dollars.

Summary statistics on the income measure are found in Table 2. The first row records the proportion of households with non-missing income data, which is 90%. The second row shows the proportion of households with zero reported taxable income, which is approximately 1%, driven mostly by female-only and male-only households. The third and fourth rows of Panel A show the extent of top-coding in the SSA data. Across all households, the average number of top-coded years is 12.19. For coupled households, the average is 15.06, while for female-only and male-only households, the average is 3.5 and 10.8, respectively, which reflects that female-only household exhibit substantially lower income. A similar patten emerges in the fourth row of Panel A, which reports the fraction of households that are never observed with a top-coded year.

The last row of Panel A of Table 2 reports the mean and standard deviation of total lifetime SSA income for the household. Average total lifetime income in 2010 dollars is about \$2,171,000 for all households and about \$2,476,000 for coupled households. Corresponding means for female-only and male-only households are \$1,301,000 and \$1,863,000, respectively. In Panel B of Table 2 we report the distribution of total SSA income by household structure. The first row reports quantiles for all households, with remaining rows doing so by household type. The tenth percentile is just over \$429,000. At the 25th percentile, household income is about \$1,122,000 in total over the lifecycle. The median is \$2,066,000 and the top 10% earned at least \$3,941,000. Patterns of differences across households are similar across the distribution, with coupled households earning the most followed by male-only and then female-only households.

3.3.3 Household Wealth

The HRS contains rich and detailed information on household wealth. Unfortunately, data related to household retirement wealth and stock market participation pose various challenges. Values of defined contribution plans from previous jobs are not asked in every wave; stock allocations in defined contribution plans are only asked in certain waves and only for plans associated with the current employer; and expected defined-benefit pension income is asked only of plans at the current employer. In some cases, such issues may be relatively unimportant. However, because this paper studies heterogeneity in wealth for elderly households, having a complete picture of retirement assets is of fundamental importance. While some data issues have no hope of being resolved, our sample comprises households for whom wealth data are most likely to be both accurate and comprehensive.

Our measure of *total wealth* is designed to encompass all components of household wealth. Our data include the present value of all pension, annuity, and social security income, which come from the RAND HRS income files, as well as the net value of housing (including primary and secondary residences as well as investment property), the net value of private businesses and vehicles, all financial assets including cash, checking accounts, savings accounts, CDs, stocks and stock mutual funds, bonds and bond mutual funds, trusts, and other financial assets, less the net value of non-housing debt. Each of these are taken from the RAND HRS wealth files.[34] Further, we include the account value of all defined contribution retirement plans.[35] We exclude from our wealth measure values of insurance.[36] All monetary values are measured in 2010 dollars. We winsorize the log of real total household wealth at the 1st, and 99th percentiles.

We note that our measure of wealth includes both marketable securities, such as stocks which can be easily sold at publicly available prices, and non-marketable assets such as social security income. Our measure of wealth is therefore intended to capture the overall financial security of households rather than the market value of household assets. Our results are qualitatively unchanged if we limit household wealth to exclude retirement income and housing, which can be interpreted as the market value of households' pure financial assets. Further details on wealth, including possible reasons for mismeasurement and alternative subsamples, are found in Appendix A.

Table 3 shows the 10th, 25th, 50th, 75th, and 90th percentiles of our various wealth measures, as well as the mean. The first row of Table 3 shows total wealth winsorized at the 1st and 99th percentiles, including both housing and the present value of all retirement income. The average for our sample is $563,060. However, the median individual has total wealth of roughly $321,040, which is substantially lower. This is due to high levels of wealth among individuals in the upper tail; the 10th percentile of wealth is $40,730, whereas wealth at the 90th percentile is $1,311,970. The second through fourth rows of Table 3 show wealth excluding the values of housing, account balances in defined contribution plans and the present value of retirement income (social security and defined-benefit pensions), and both housing and retirement income. A few interesting patterns emerge. First, housing makes up a larger portion of total wealth at the lower end of the distribution. For example, at the 10th percentile housing wealth is nearly half of total wealth, whereas it accounts for a little less than one-fourth of total wealth at the 90th percentile. A similar result applies to

[34] When calculating the present discounted value of annuity, social security, and defined-benefit pension income, we follow Yogo (2016) and assume a 1.5% guaranteed rate of return, discounted by the probability of death in each year conditional on age, cohort and gender of the financial respondent as determined by the Social Security life tables.

[35] Plans that are maintained either at previous employers for working households, or are still maintained by the previous employer for retired households, are referred to by the HRS as "dormant plans."

[36] Without further details on the structure or terms of specific insurance products it is difficult to estimate a market value for these items.

retirement wealth. In fact, for individuals at the 10^{th} percentile, housing and retirement wealth comprise the entirety of household wealth.[37]

4 The EA Score and Wealth

4.1 EA Score in the HRS Sample

Figure 1 plots the smoothed distribution of the EA score for our analytic sample. The EA score is normalized to have mean zero and variance of one, and is approximately normally distributed.

The first four columns of Panel A of Table 4 report average values of our basic demographic variables for each quartile of the individual EA score distribution. Note that we have non-missing EA scores for 5,692 distinct individuals and here we divide them into quartiles based on their individual EA scores, not the average EA score in their household. Column [5] reports p-values from a test of mean differences comparing the fourth and first quartiles. Individuals in the highest EA score quartile tend to be older, are more likely to be male, have higher values for both mother's and father's years of schooling, have more years of schooling themselves, and are more likely to be members of a two-person household.

Importantly, because the SNPs associated with the EA score are not found on sex chromosomes, the slightly higher representation of men in the fourth quartile of the EA score must result from selection.[38] This is likely because higher EA score people live longer on average (see Section 5 below), and men have higher average mortality rates than women. Combined with the fact that one must have survived until at least 2006 to be genotyped, it is unsurprising that higher EA score people tend to be slightly over-represented by men.[39]

Panel B of Table 4 shifts attention to the household and divides households into quartiles based on the average EA score. The panel reports average values of demographic variables for men and women separately. The same patterns emerge in Panel B as in Panel A; higher EA score households tend to consist of younger, more educated individuals with more educated parents. Women tend to be younger than men, and have slightly lower years of schooling.

[37]In Appendix A, we show summary statistics for wealth at a more disaggregated set of sources, including IRAs, stock holdings, cash, and CDs. For wealthier households, wealth is further diversified, including items such as secondary homes and real estate.

[38]Recall from our discussion in Section 2 that the polygenic score does not reflect variants on sex chromosomes, so its distribution is identical in men and women.

[39]To explicitly test whether this is true, in results available from the authors, we show that restricting attention to individuals born during or after 1928 (the first quartile of birth year in the sample) eliminates gender differences in EA score, suggesting that differences are driven by older birth cohorts, where survival selection plays a role.

The gender differences apparent in Table 4 raise the more general concern of selection into being genotyped. The 5,692 genotyped individuals in our sample comprise roughly 15 percent of HRS respondents. In Appendix A, we compare genotyped and non-genotyped individuals. We find substantial differences, particularly in education, income, and wealth. These differences raise the possibility that our results might be different among individuals choosing not to be genotyped. We can only speculate about the external validity of our results, though several empirical patterns are reassuring in this regard. First, as additional individuals in the HRS are genotyped our results remain largely unchanged.[40] Second, the relationship between the EA score and education is robust across data sets (Okbay et al., 2016). Third, the relationship between education and wealth is similar in the genotyped and the non-genotyped sample.[41]

Because the household is our main unit of observation, another important consideration is potential assortative mating on individuals' EA scores. For example, individuals with high EA scores may be in households with other high EA score individuals. This could arise due to similarities in schools, neighborhoods, or other environments where social activity is correlated with socioeconomic status, or because high EA score individuals are more likely to meet their partner through post-secondary education.

We explore the extent of assortative mating in Table 5, where we restrict attention to the 1,477 households in which both members have non-missing EA scores. First, we calculate the quartiles of the distribution of individual EA scores for all individuals in these households. Each column focuses on women in a different quartile of the individual EA score distribution, with the rows indicating the distribution of their partners' scores across the quartiles of the EA distribution.[42] Panel A reports these distributions for individuals sorted by the raw, unadjusted value of their individual EA scores; Panel B reports distributions where individuals have been sorted based on the residual in a regression of their individual EA score on degree dummies and years of schooling. With perfect assortative mating, the matrices reported in Table 6 would be diagonal matrices, with 100% populating the diagonal entries and 0% populating the off-diagonal entries. Alternatively, random assignment would generate matrices with 25% for each entry.

We find some evidence of assortative mating, especially among the highest and lowest

[40]In Appendix B, we compare results from an older version of the EA score with and without information from HRS respondents who were genotyped in 2010. Results do not change when we add individuals genotyped in 2010.

[41]In Appendix B, we also show robustness of main results to the use of HRS sampling weights, which provides some evidence that findings are not driven by selection on variables in the HRS not included in our analyses.

[42]This exercise closely follows Charles, Hurst, and Killewald (2013) (see their Table 5 on p. 61), who examine assortative mating on parents' wealth.

EA score quartiles. We find that 29.2% of females in the lowest individual EA score quartile are coupled with males in the lowest individual EA score quartile, compared to 28.2% with males in the second quartile, 24.7% with males in the third quartile, and 18.0% with males in the fourth quartile. Columns for women in the second and third quartiles of the individual EA score distribution show less assortative mating, with values ranging from 22-27%. The column corresponding to women in the fourth quartile of the score distribution once again shows stronger evidence of assortative mating, with only 20.3% of the highest EA quartile females coupled with the lowest EA-quartile males, compared to 32.9% with the highest EA quartile males. Although we are able to reject the random-assignment null hypothesis that all entries are equal to 25% ($p < 0.001$), the degree of assortative mating appears modest relative to the counterfactual of perfect sorting. Indeed, while the within-couple correlation of years of schooling is around 0.51, the within-couple correlation of the individual EA score is only around 0.13. Moreover, in Panel B, where we adjust for highest degree attained and years of education, we find less evidence of assortative mating and fail to reject the null hypothesis of random sorting ($p = 0.21$).

Because individuals' EA scores appear to vary considerably within two-person households, our analyses relating average household EA score to household wealth may miss important dimensions of decision-making by using the average score as our primary measure. For example, if the highest EA score individual in a household is the primary decision maker, our results may be attenuated by using the average score, as a household including two moderate-score individuals will be treated the same as a low-high score household. For this reason, we provide a host of robustness checks based on alternative specifications of the household EA score. These are shown in Appendix B. All of our main results are robust to these alternative specifications.

4.2 The EA Score and Wealth

4.2.1 Main Association

Figure 2 presents the main association we study in this paper. The top panel plots the unconditional, nonparametric (Lowess) relationship between the log of total household wealth and the average household EA score. The results are striking. The relationship between the EA score and wealth is increasing for normalized values of the EA score between -2 and 1 (over 80% of the sample), although it flattens and even declines somewhat after an EA score of 1. The size of the wealth differences are economically large; moving from an EA score of -1 to 1 implies a percentage change in wealth of approximately 72%, or the equivalent of just under $196,000.

In the second and third panels of Figure 2, we examine whether the economically significant relationship between the EA score and wealth holds in different subsamples. In the second panel, we look only at households where all members have a high school degree or less, and plot the relationship between wealth and the EA score for all wealth-year observations for two separate time periods: 1992-2006 and 2008-2012, which comprise years before and after the financial crisis. The third panel repeats this exercise for households where at least one member has a college degree or more. In each of the four subsamples, the relationship between the EA score and wealth is positive and substantial for EA scores between -2 and 1. For values of the EA score greater than 1 the relationship becomes flat (or even slightly negative).[43]

Figure 2 offers compelling evidence that the EA score and wealth are positively associated. We examine this relationship more formally in Table 6, which reports results from regressing log household wealth on the EA score for specifications with various sets of controls. Standard errors are clustered at the family level.[44] Column [1] shows the unconditional relationship between the EA score and the log of household wealth with no additional covariates. A one standard deviation increase in the EA score is associated with 33.1% greater wealth. This result is highly statistically significant, with a t-statistic greater than 14. In Column [2], we add basic controls for age (separately for males and females), birth year (separately for males and females), sex of the financial respondent, calendar time, and family structure.[45] Henceforth, including in tables, we refer to this set as "standard controls." The inclusion of standard controls has only a marginal effect on the coefficient on the EA score, which remains large and highly significant. In Column [3], we include the first 10 principal components of the genetic data separately for the male and female household members.[46] These variables

[43]One possible concern is that households with relatively more observations, e.g., due to low mortality rates, exhibit different relationships between the EA score and wealth and are over-represented in these figures and in subsequent analysis. In Appendix B, we show that main results are unchanged if we only use one observation per household.

[44]Multiple households could be linked in our data if a once-married couple divorces or separates to become two distinct households. In such a case, the individuals in the divorced household would belong to three distinct households in our data, but just one family.

[45]Specifically, we add the following: a set of dummies for every possible age for the male household member, interacted with an indicator for a male only household, a complete set of dummies for every possible age for the female household member, interacted with an indicator for female only households, complete sets of dummies for male and female birth years, also interacted with indicators for male only and female only households respectively, dummies for calendar year, an indicator for male financial respondent, and dummies for a male only household and female only household.

[46]Specifically, we include the first 10 principal components for the male household respondent, along with interactions with a dummy for being in a male only household, the first 10 principal components for the female household, along with interactions with a dummy variable for being in a female only household, separate dummies indicating missing genetic data for the male and female household members, respectively, an interaction between the missing male genetic data indicator and a dummy for a male only household, and an interaction between the missing female genetic data indicator and a dummy for female only household.

are intended to approximate family fixed-effects as explained in Section 2 (Benjamin et al., 2012). The principal components reduce the EA score coefficient from 0.319 to 0.305, and it remains statistically significant.

In Column [4] of Table 6 we add controls for years of schooling for each member of the household. Including years of schooling significantly reduces the size of the gene-wealth gradient, decreasing the coefficient to 0.141. This is unsurprising; the EA score was developed based on years of schooling, and education undoubtedly affects income and wealth accumulation over the life cycle. It is important to note, however, that the coefficient remains economically important and statistically significant even after controlling for years of schooling. A coefficient of 0.141 suggests a one standard deviation increase in the genetic score is associated with 14.1% greater wealth during retirement. In Column [5], we include more flexible measures of education. Instead of the simple count of years of schooling for each member, we include the following: a complete set of dummy variables for every possible number of years of schooling for the male household member, a complete set of dummies for every possible completed degree for the male household member, interactions between all male education dummies and an indicator for male-only households, a complete set of years of schooling dummies for the female household member, a complete set of degree dummies for the female household member, interactions between all female dummies and an indicator for female-only households, a full set of interactions between the male and female years of schooling dummies, and a full set of interactions between the male and female degree dummies. We refer to this set as "full education controls" in our subsequent analyses, including in tables reporting estimates. Including the full set of education controls reduces the EA score coefficient to 0.124. Even in this specification the coefficient remains highly statistically significant.

In Column [6], we exclude the education controls included in Columns [4] and [5] and add controls for labor income. In particular, we include the total of lifetime earnings for the household from the SSA data described in Section 3. Controlling for income reduces the coefficient on the EA score from 0.305 to 0.291, which remains statistically significant. In Column [7], we add the full set of education variables along with income and other controls. The results are consistent with Columns [5] and [6]. The coefficient on the EA score is 0.115, (p-value $< .00001$), suggesting that a one standard deviation increase in the EA score is associated with 11.5% greater wealth.

Table 6 indicates that the EA score is associated with wealth even after controlling flexibly for completed schooling and degree type. One interpretation of this result is that the score measures biological traits that promote wealth independently of any effects on the acquisition

The principal components for individuals who are not genotyped are set to zero.

of human capital. However, it could also be the case that the education variables in the HRS are measured with error, or do not fully reflect the educational investments associated with genetic factors. If this is true, then the remaining genetic gradient in Column [7] may simply stem from the effects of unobserved human capital investments rather than biological traits. In particular, our control set does not include measures of school quality, which has been studied as a potentially important dimension of educational investment (Behrman and Birdsall, 1983).[47]

Measurement error in the HRS education variables could explain some of the remaining coefficient on the EA score. However, given results linking higher quality teachers to higher adult earnings (Chetty, Friedman, and Rockoff, 2014), we would reasonably expect that if unobserved school quality were playing a major role, controlling for lifetime earnings should substantially alter the relationship between the EA score and wealth. Our results suggest that controlling for lifetime income does little to reduce the genetic gradient in wealth. Consistent with our results, Smith-Woolley et al. (2018) fail to find evidence of a relationship between school quality and national exam results in the U.K. after controlling for an earlier version of the EA score.

Nonetheless, measurement error in income is still a concern. It may be that complete measures of income that do not suffer from top-coding or reporting biases fully account for the gene-wealth gradient once education (even improperly measured) is included. Unfortunately, there is no perfect way to address these issues. One approach we pursue is to include self-reported contemporaneous labor income from the HRS as an additional measure of household income. Results from this exercise are discussed in greater detail below. We find that even after controlling for HRS income, or both HRS and SSA income data simultaneously, the gene-wealth gradient remains economically large and statistically significant. Regardless, the reader should interpret our results with these potential measurement issues in mind. These issues may be particularly salient when we address other potential mechanisms linking the EA score to wealth in Section 5. For this reason, the results in Section 5 should be viewed only as suggestive.

[47]Recent evidence on school quality is mixed. Some papers show evidence that charter schools and schools with more funding improve outcomes on test scores and post-secondary educational outcomes (Deming et al., 2014; Jackson, Johnson, and Persico, 2015; Angrist et al., 2016) and reducing racial achievement gaps (Dobbie and Fryer Jr, 2011). Other work shows that the impact of higher school quality is very small once selection into more prestigious schools is accounted for (Abdulkadiroğlu, Angrist, and Pathak, 2014). See Card and Krueger (1996) for a survey of earlier literature on school quality effects.

4.2.2 Robustness

Figure 2 and Table 6 show a strong, economically large relationship between the average household EA score and household wealth. In Table 7, we provide results from three alternative specifications with potentially important differences. Each column includes the same controls as the corresponding column in Table 6. An expanded set of robustness tests are found in Appendix B. In Panel A, we replace the average household EA score with separate individual scores for the financial respondent (FR) and non-financial respondent (NFR). The financial respondent answers financial questions on behalf of the household. If, for example, the financial respondent has sole responsibility for the financial decisions of the household, this respondent's EA score may have a larger association with wealth accumulation than the household average score. It may also be that conditional on the financial respondent's EA score, a higher EA score of the non-financial respondent could also associate with greater wealth, a result that would also be obscured by the average.[48] The average individual EA score for the FR is 0.07, while for the non-FR it is -0.01. About 65 percent of financial respondents are male. These differences suggest two possibly countervailing effects. First, some households delegate financial decisions to the member with a higher EA score. Second, in some households the male makes financial decisions despite having a lower EA score.

The first three columns of Panel A in Table 7 show that the coefficient on the individual EA score for the FR is substantially larger than the coefficient on the NFR score (0.255 vs. 0.156), though both are statistically significant. Once we control for education in Columns [4] and [5], the size of the coefficient on the NFR score shrinks and it is only marginally significant. When income is included (but not years of schooling) in Column [6], the coefficients remain similar to the estimates in Column [3]. In Column [7], we include both income and our full set of education controls, and find that the coefficient on the FR EA score remains large and statistically significant, while the coefficient on the non-FR score again shrinks and is only significant at the 10% level. In general, these results suggest that the FR score plays a more important role than that of the NFR, though both contain explanatory information.

In Panel B of Table 7, we re-estimate all seven specifications while including both retired and non-retired households. For non-retired households with defined-benefit pensions, economic resources are understated since we do not include expectations of future defined-benefit income. Compared to individuals in our main analytic sample, individuals added for this analysis are younger (financial respondents have an average age of 73 in the baseline sample as opposed to 61 in the added household-year observations). The households that are

[48]This could occur if partners exchange information, a point made in Benham (1974) who studies the benefits of women's education for the household.

added to this sample tend to be more highly educated than the baseline (by roughly a full year of schooling for both men and women) and exhibit higher lifetime income ($2,695,255 versus $2,122,583 for our baseline sample). Estimates in Panel B are highly similar to the results from our main specifications in Table 6, which suggests our restriction to retired households is not an important factor driving the relationship between the EA score and wealth. We note that in Column [7], the coefficient on the EA score in Panel B is 0.104 compared to 0.115 in Column [7] of Table 6, so including non-retired households reduces the coefficient slightly, but the coefficient remains highly significant.

Finally, in Panel C of Table 7, we include as an additional control the log of the average of the household's self-reported labor income streams from the HRS.[49] In this specification we include only the years in which the household has at least one working member. The self-reported income data in the HRS avoids the top-coding issues of the SSA data. However, because income in the HRS is only reported for years in which the household is working, and because the HRS is a sample of elderly Americans, this necessarily means that HRS labor income is observed toward the end of the life-cycle or not at all. These differences are meaningful. Average annual household income using HRS data is $62,535 and the correlation coefficient between the log of this HRS average and the log of total income using SSA data is 0.31. Indeed, Panel C reveals that both measures of income independently predict wealth. Nevertheless, the estimated coefficient on the EA score is 0.108 when both income measures are included — virtually identical to our baseline estimate.[50]

5 Mechanisms

This section considers possible channels beyond income and education through which the EA score may relate to wealth. The evidence in this section is only suggestive, and may highlight potential areas for future research. Specifically, we investigate parental education and inheritances, mortality and savings rates, and investment decisions such as stock market participation, home ownership, and business ownership. We also consider how the EA score relates to different dimensions of financial decision-making, including beliefs about the stock market and reported planning horizons. Finally, we show differences in the gene-wealth gradient depending on whether individuals receive income from defined-benefit pensions.

[49]Specifically, for each member of the household, we consider only years in which they are not retired and report working for pay. We add up real income for each observation within a particular household-year, and average across available years in the HRS up through 2010.

[50]In Appendix B, we estimate additional specifications which include various quantiles of the income data along with a specification that only uses HRS income data. In each specification, we find that our results remain quantitatively unchanged.

Summary statistics for each potential mechanism are provided in Table 8. Similarly to Table 1, we report means and standard deviations for all households and for different household structures separately. We discuss these variables when analyzing each potential mechanism below.

5.1 Transfers and Parental Education

First, we investigate whether higher EA scores predict higher wealth due to transfers from parents. This is a particularly important channel given that children also inherit their genes from their parents. Summary statistics on the relevant variables are found in Panel A of Table 8, which includes an indicator for whether the household ever receives an inheritance, real total inheritance amounts among those who receive one, and the within-household average of parental years of schooling. Among other things, parental education proxies for the resources parents could provide to their children. About 34% of households report direct transfers in the form of an inheritance and, among those who do, the average amount is $143,200 (2010 dollars). The average level of fathers' education is 9.6 years of schooling, while the average level of mothers' education is 10.1 years.

In Table 9, we relate the EA score to inheritances. All regressions include our standard controls and full education controls, unless otherwise noted. In Column [1] of Table 9, we estimated a cross-sectional regression where the dependent variable is an indicator variable equal to one if the household has ever received an inheritance over the span of the sample. The estimated coefficient on the EA score suggests that a one standard deviation increase is associated with a 2.1 percentage point increase in the probability of receiving an inheritance. In Column [2] we estimate a cross-sectional regression where the dependent variable is the log of the real dollar value of all inheritances received over the sample. Because the log of total inheritances is defined only for values greater than zero, this specification is equivalent to a regression of inheritance values conditional on receiving an inheritance. We find no relationship between the EA score and the size of inheritance wealth conditional on receiving an inheritance. This suggests that the EA score is related to inheritances at the extensive margin but not the intensive margin.

Next, we regress different measures of parental education on the household's average EA score. In Column [3], the dependent variable is the average education of the fathers of both household members, and we include our standard controls but no measures of respondent education. Results suggest that a one standard deviation increase in the EA score is associated with over 0.84 more years of average fathers' schooling, and the coefficient is highly statistically significant. Column [4] presents the same specification but with the average of mothers' education as the dependent variable. The coefficient suggests an increase in 0.66 years of schooling of mothers for one standard deviation higher household EA scores. In columns [5] and [6], we investigate whether the relationship between parental education and the EA score is entirely explained by household members' own education. In Column [5] the dependent variable is again average fathers' education, but we now include the full set of household education controls. Column [6] reports analogous coefficients with average mothers' education as the dependent variable. The estimated coefficients on the EA score are reduced dramatically in these specifications, from 0.84 to 0.33 for fathers' education and from 0.66 to 0.19 for mothers' education. This is unsurprising given the intergenerational persistence of education (Black, Devereux, and Salvanes, 2005). However, the coefficients remain statistically significant, which indicates that household environments and other investments could play a role in wealth accumulation beyond just educational attainment.

In summary, the results in Table 9 suggest that the EA score is highly correlated with parents' education, although the magnitude of this association decreases substantially when household members' own education is included. Further, a higher EA score does appear to be associated with a greater likelihood of receiving an inheritance, although the size of the inheritance conditional on receiving one is not related to the score.

5.2 Mortality and Savings

Another way in which wealth may be related to genetic endowments is through longevity. If individuals with higher individual EA scores expect to live longer, they may endogenously save more to finance these additional years of consumption. Furthermore, longer expected lives may lead to longer investment horizons, which may affect the mix of assets in household portfolios. Panel B of Table 8 contains summary statistics on annual mortality rates for years after being genotyped and the subjective expectation of living to 75 years old, for individuals in our analytic sample.

We begin our analyses by directly estimating the empirical relationship between the individual's EA score and mortality. We construct an indicator variable equal to one if the individual dies in the next year, and estimate a linear probability model of the likelihood of dying in a particular year as a function of the individual's own personal EA score, the principal components, and dummy variables for age, birth year, years of schooling, and degree. We restrict this regression to person-years in which an individual was between the ages of 50 and 90, and we drop years before an individual was genotyped. Panel A of Table 10 provides the results of this regression. In Column [1] we include both females and males in the sample, and find that the individual's EA score is associated with a 0.3 percentage point decline in mortality. Columns [2] and [3] consider females and males, respectively. The estimated association for females is the same as in the pooled sample, but it is both smaller and statistically indistinguishable from zero for males.

We also consider beliefs about mortality. In principle, objective mortality should only affect behavior if individuals expect to live longer. In this sense, beliefs about mortality are perhaps the more relevant mechanism linking genetic endowments to wealth. The HRS repeatedly asks individuals to provide their subjective beliefs for the probability that they will live to the age of 75. In Column [4], we regress this subjective belief on the individual EA score, our standard controls, and the full set of education controls in a sample of individuals aged 50-65. We do not find a significant association between the EA score and the level of this subjective probability. We also estimate this regression for females and males separately in Columns [5] and [6], and find that for females a one standard deviation rise is the individual's EA score predicts a 0.66 percentage point rise in reported beliefs about living to age 75. For males, the relationship is negative and statistically insignificant.[51]

Next, we explore whether the average household EA score predicts higher savings rates. One way in which the EA score may relate to wealth is through greater precautionary saving, either due to lower expected mortality as discussed earlier, or due to better financial planning. Unfortunately, the nature of the HRS data makes this a difficult exercise. Ideally, we would like to observe household savings rates over the life-cycle, particularly during the early years in which savings benefit the most from compounded interest. However, the HRS contains individuals who are either at the very end of their working lives or have already retired. A further complication is that such households are likely to consume out of accumulated wealth rather than labor earnings, making standard measures of savings rates less informative.

In Panel B of Table 10 we estimate whether the EA score is empirically related to spending in the HRS. Our measure of spending comes from the RAND files derived from the Consumption and Activities Mail Survey (CAMS), collected by the HRS in off-wave years beginning in 2001 and which include data on household expenditures on nondurables, durables, housing, and transportation.[52] We include only households that have at least one member still working (not retired), and aggregate SSA labor income in that year at the household level. Column [1] shows results from a regression of total household expenditures on indicator variables for each labor income quintile in addition to the EA score and our standard controls. In Column [2], we restrict to households in which both members are less than 60 years old

[51]The important question of whether the EA score predicts incorrect beliefs about mortality risk, and whether such deviations could differ by gender, is left for future study.

[52]In results available from the authors, we show that our results are unchanged if we use the consumption variable in the RAND HRS data, which attempts to measure the consumption flow of spending on durable goods rather than the simple dollar amount of expenditures.

in order to capture households that are likely still in the high-earning years of their careers. In both Columns [1] and [2], households in the highest income quintile spend considerably more than households who earn less. However, in each case the coefficient on the EA score is small and statistically insignificant, which suggests that after controlling for earnings, higher EA score households spend the same amount as low EA score households.

One concern may be that, despite the restriction to households with all members less than 60, we are still capturing households with low labor earnings who are consuming primarily out of wealth. We first address this concern in Column [3] by dropping households in the first quintile of income (less than or equal to $11,732 in 2010 dollars). In this specification, the coefficient on the highest income quintile becomes smaller and less statistically significant, but the coefficient on the EA score is unchanged from Column [2] and remains statistically insignificant. To more explicitly address the concern that individuals in their later working lives consume primarily out of savings rather than labor income, in Columns [4]-[6] we repeat the specifications in [1]-[3], but include wealth in the previous wave as an additional control. In each specification, the coefficient on wealth is highly statistically significant and economically large. In each, the coefficient on the EA score drops in magnitude and remains statistically insignificant.

Results in Table 10 show that, while both expected and realized mortality are related to the individual EA score, measures of spending are unrelated to the EA score conditional on income or wealth. This raises doubts about their potential importance as a mechanism linking the EA score and wealth. This finding is consistent with results using twins data that suggest the genes implicated in savings are not related to education (but are negatively related to genes implicated in smoking and obesity, suggesting a relationship between self control and savings (Cronqvist and Siegel, 2015)). However, an important caveat to our findings is that the expenditure data in the HRS have a variety of shortcomings. A stronger conclusion about savings as a potential mechanism would require a better test of the relationship between savings and the EA score from data on younger households.

5.3 Risk Aversion

While we find no empirical evidence that the EA score is related to the level of savings, we next examine if the EA score is associated with differences in *how* households save. A well-established source of heterogeneity in household wealth are returns to risky endeavors, such as participation in risky asset markets or business ownership. One mechanism that may therefore relate the EA score to wealth is aversion to risk. To examine the relationship between risk aversion and the EA score, we use questions in the HRS designed to elicit measures of risk tolerance based on hypothetical income and wealth gambles. Generally, these questions pose hypothetical scenarios in which the respondent faces a choice between a guaranteed endowment of wealth or stream of income, or a 50-50 gamble that will result in a permanent increase or decrease in that endowment or income. Specifically, respondents are asked to choose between two jobs: "The first would guarantee your current total family income for life. The second is possibly better paying, but the income is also less certain. There is a 50-50 chance the second job would double your total lifetime income and a 50-50 chance that it would cut it by X." The series replaces X with a set of possible income losses: ten percent, twenty percent, one-third, one-half, seventy five percent. Additionally, respondents are asked one of two hypothetical wealth gambles with a similar structure. One is based on an inherited business worth one million dollars today, or that may be sold in one month with a 50-50 chance of being worth two million dollars or X. The other is based on an immediate inheritance worth one million dollars, with the potential to participate in a risky business venture that has a 50-50 chance of doubling in value or falling in value by X. In each case, X varies by the same proportions as the hypothetical income gamble.

Based on the responses to these hypothetical gambles, each respondent can be grouped by the smallest downside for which they still reject the gamble. We also create a dummy variable for each gamble that takes a value of one if an individual always responds with a preference for the guaranteed wealth or income.[53] A value equal to one for this variable indicates the highest degree of risk aversion permitted with this set of questions. Summary statistics on the risk aversion parameters are found in Panel C of Table 8.[54] According to Table 8, 39% of respondents would not take a gamble that would double their income with 50% probability and cut it by 10% with 50%. These are the individuals assigned to the highest degree of risk aversion for this question. On the other extreme, 5% of respondents would take a 50-50 gamble where the downside is a 75% reduction in income.

In Column [1] of Table 11, the dependent variable is our binary indicator for the highest degree of risk aversion based on the labor income gamble. We find a negative association between the average household EA score and risk aversion — a one standard deviation increase in the score is associated with a reduction in the probability of the most risk averse response by 2.2 percentage points. In Columns [2] and [3], we use indicators for least willingness to take the gamble using the inheritance and business risk questions to construct the dependent variables. We find no statistically significant relationship between the EA score and risk aversion for the inheritance question, but do find that the probability of a respondent

[53] One possibility for future research is to follow Kimball, Sahm, and Shapiro (2009), who suggest ways to improve measures of risk aversion in large data sets with noisy measures.

[54] We only include answers for the income gambles. Summary statistics for variables from the two other risk aversion questions are found in in Appendix A.

giving the most risk averse response for the business risk question is 2.6 percentage points lower for a one standard deviation increase in the EA score, which is significant at the 0.05 level.

In Columns [4]-[6], we allow the outcome variable to be an ordered categorical variable indicating the riskiest gamble that a respondent accepts. This variables can take one of six values, with higher values corresponding to higher degrees of risk aversion. We estimate an ordered probit model in these specifications, and report coefficients for the latent index. Column [4] shows that the EA score is associated with a significant decrease in the latent index for risk aversion. Columns [5] and [6] repeat the ordered probit estimation for the inheritance and business wealth gambles, respectively. Again, we find no statistically significant relationship between the EA score and risk aversion based on the hypothetical inheritance wealth gambles, but do find a significant relationship with risk aversion for the business wealth gamble.

5.4 Stocks, Housing, and Business Ownership

Continuing with our examination of how households save, we next examine whether the EA score is related to stock market participation, business ownership, and owning a home. Each of these asset classes is the subject of a well-established literature highlighting their importance as a source of heterogeneity in wealth accumulation over the life-cycle. Panel D of Table 8 provides summary statistics on these portfolio decisions. For example, 82% of households own a house, while 8% own a business and 44% own stocks.

Panel A of Table 12 regresses indicator variables for stock market participation, business ownership, and home ownership on the average household EA score and our full set of standard controls, including education variables. Columns [1]-[3] also include the log of total lifetime household income from the SSA data as an additional control. In Column [1] we find no statistically significant relationship between home ownership and the EA score, but we do find a significant relationship between home ownership and lifetime earnings. In Column [2] of Panel A, the dependent variable is business ownership, and we find a positive but statistically marginal relationship with the EA score, and no relationship between business ownership and lifetime labor income. Column [3] shows a strong positive association between the EA score and stock market participation. A one standard deviation increase in the EA score is associated with a 4.8 percentage point increase in the probability of owning stocks, and this coefficient is statistically significant at the 1% level. Compared to an average rate of stock ownership of 44%, the coefficient suggests this predicted increase in participation is also economically meaningful.

Of course, stock market participation and business ownership are likely affected by accumulated wealth, and we have already found the EA score is associated with wealth. This means that the relationships between the EA score and stock market participation and business ownership could be operating purely through wealth (Charles and Hurst, 2003). We address this in Columns [4]-[6], which repeat the specifications in Columns [1]-[3] but also include the log of financial wealth from the previous wave. Consistent with the existing literature, we find that the coefficient on lagged wealth is large and statistically significant in all three specifications. For home ownership, the coefficient is now negative and statistically significant, which is not the expected result. It suggests that conditional on lagged wealth, a higher EA score is associated with a lower likelihood of home ownership, which could reflect a higher probability of divesting homes after retirement among higher score households. We find no evidence for a relationship between the EA score and business ownership after controlling for lagged wealth. However, the relationship between the EA score and stock ownership remains significant and economically meaningful. A one standard deviation increase in the EA score is associated with a 3.5 percentage point higher likelihood of owning stocks, with statistical significance at the 0.01 level. Because stocks have traditionally offered substantially higher returns than other liquid securities such as money-market funds or bonds, this may be an important factor for explaining the gene-wealth gradient, and may also suggest that these genetic endowments provide a microfoundation for the persistent differences in returns to wealth.

To examine the extent to which the important components of household saving — home, business, and stock ownership — can be possible explanations for the association between the EA score and wealth, we include each as controls in regressions of wealth on the EA score and our standard controls. Column [1] of Panel B in Table 12 establishes the baseline coefficient by repeating the final specification in Table 6, but restricting the sample to those households with non-missing values of the portfolio and parental variables. In Column [2] of Panel B, we also include controls for various measures of parental transfers, including an indicator for whether the household has ever received an inheritance, the log of the sum of all inheritances received (with this value set to zero for households not receiving inheritances), and controls for mothers' and fathers' education. In this specification, the coefficient on the EA score decreases from 0.124 to 0.114, but remains statistically significant. In Columns [3] and [4], we include indicator variables for whether the household owns their home or has ever owned a business during the sample. In each case, the coefficient on the EA score decreases slightly, to 0.117 and 0.116, respectively, but remains statistically significant. In Column [5], we include an indicator for stock ownership. This reduces the coefficient on the EA score substantially, from 0.124 to 0.072, a reduction of roughly 40%.

Finally, in Column [6], we include controls for all four potential mechanisms. Together, these controls reduce the coefficient on the EA score to 0.066. In other words, the full set of investment controls — home, business, and stock ownership — reduce the coefficient to 0.066 from 0.114 (from the specification that includes only parental transfers). This offers preliminary evidence that investment decisions over the life-cycle, broadly defined, may be an important mediator of the gene-wealth gradient.

However, we caution that these results should be interpreted with care. The mediation exercise performed in Panel B of Table 12 may be biased if measurement error in the right-hand-side variables leads to the incorrect interpretation of the coefficients. We reiterate that the results in this section are suggestive of possible mechanisms, but are not definitive. Motivated by these findings, in the next section we evaluate the extent to which the EA score is related to financial decision-making. In particular, we study the empirical relationships among the EA score, beliefs about macroeconomic events and the financial planning horizon.

5.5 Extreme Beliefs and Planning Horizons

An important element of financial decision-making is an assessment of the risks and uncertainties associated with the macroeconomy and the payoffs to alternative financial choices. Yet, inferring the likelihood of uncertain events can be difficult. Despite the typical assumption of rational expectations, it has long been recognized that individuals may have trouble forming accurate beliefs about probabilistic outcomes (Savage, 1954; Kahneman and Tversky, 1972). Further, a well-documented challenge for prudent savings and investment decisions is the complexity associated with intertemporal choices. Thinking about the distant future is difficult; as the planning horizon increases so too does the uncertainty around financial needs, investment and employment opportunities, family composition, and a host of other important considerations. In this section, we evaluate whether the EA score is associated with an aptitude for abstract and complicated financial decisions.[55]

Recent literature examines the role of subjective expectations in economic decisions such as human capital investments (Wiswall and Zafar, 2015) and stock market participation (Arrondel, Calvo Pardo, and Tas, 2014). Another set of papers demonstrates links between subjective beliefs and investment behaviors that impact household wealth (Lillard and Willis,

[55]An alternative would be to use the cognitive test score administered to HRS respondents to evaluate whether the gene-wealth gradient works through cognition. In Appendix B, we include the test score as an additional control and find that it does little to mediate the gene-wealth gradient. This is not surprising given why and how the test is constructed: to capture cognitive decline by asking very simple and factual questions. As discussed earlier, the correlation between the cognitive test score and the EA score is roughly 0.20 (Papageorge and Thom, 2016). Further, our results on macroeconomic beliefs remain after controlling for the cognitive test score.

2001; Dominitz and Manski, 2007; Hudomiet, Kézdi, and Willis, 2011).[56] Lumsdaine and Potter van Loon (2017) study differences in how individuals report beliefs about stock market returns, arguing that their findings reflect heterogeneity in individuals' understanding of the laws of probability. In related work Lusardi, Michaud, and Mitchell (2017) demonstrate that heterogeneity in returns to savings, which are plausibly determined by financial knowledge, can explain a substantial proportion of wealth inequality.

We begin this section by investigating whether the average household EA score is associated with differences in beliefs about macroeconomic events that are relevant for financial choices. The HRS data are uniquely well-suited for this analysis, as most respondents are repeatedly asked to provide subjective probabilities of a range of events. Individuals are asked to provide a probability on a scale of 0 to 100 for the following three macroeconomic events:

- **Stock Market Goes Up**: "By next year at this time, what is the percent chance that mutual fund shares invested in blue chip stocks like those in the Dow Jones Industrial Average will be worth more than they are today?"
- **Economic Depression**: "What do you think are the chances that the U.S.economy will experience a major depression sometime during the next 10 years or so?"
- **Double Digit Inflation**: "And how about the chances that the U.S. economy will experience double- digit inflation sometime during the next 10 years or so?"

First, we construct one (of possibly many) measures of "objectively correct" responses to these questions. Our objective benchmark probability for the stock market going up in a single year is 74 percent, which is the probability the S&P 500 increases in value in a given year for the period 1963-2010. There is no common definition of an economic depression, but clearly this refers to an unusually severe period of economic contraction. We use data from the Federal Reserve Bank of Saint Louis on annual real GDP growth over the period 1948-2016, and define an unusually severe contraction as a year with growth less than or equal to -0.73 percent, which is the 25th percentile of the distribution of growth rates for negative-growth years. Based on this metric, the unconditional probability of a severe contraction is 4.4 percent per year, which implies a 36 percent probability for such an event over a 10 year period. Finally, the Bureau of Labor Statistics reports two years with

[56] Hurd (2009) provides a review of subjective probabilities reported in household surveys such as the HRS. A number of researchers have used the HRS to study cognition, probabilistic thinking and investment decisions (Lillard and Willis, 2001; Kézdi and Willis, 2009, 2003). Another set of related studies focuses on cognitive decline and retirement decisions (Rohwedder and Willis, 2010; Kézdi and Willis, 2013; Delavande et al., 2006; Delavande, Rohwedder, and Willis, 2008).

double digit inflation (1980, 1981) over the period 1958-2015. This implies an approximate probability of 3.4 percent for double digit inflation in any year, or about a 29 percent chance for double digit inflation over a 10 year period.[57] In Panel E of Table 8, we provide summary statistics for expectations about stock market performance, including raw reported probabilities, deviations from objective probabilities, and heaping on focal probabilities.

Panel A of Table 13 provides estimates of the association between the average household EA score and individual beliefs about the probabilities of these macroeconomic events. We use the average household score rather than the individual EA score so as to be consistent with our analysis in Section 5.4, and to avoid decisions about intra-household information transfers. Our first measure is the absolute value of the deviation between the respondent's subjective probability and the objective probability. We regress this deviation on our standard controls and the EA score in Column [1]. For all three events, higher values of the polygenic score are associated with a statistically significant reduction in the deviation between the respondent's subjective probability and the objective probability. For example, in Column [1] of Panel A, the coefficient estimate of -0.572 suggests that a one standard deviation increase in the EA score is associated with a reduction in the deviation from the objective stock market increase probability of over one half of one percentage point. Coefficients of -0.537 and -1.047 are estimated for the depression and double-digit inflation questions, respectively.

Columns [2]-[4] of Panel A in Table 13 examine binary outcomes indicating whether respondents answered with specific focal probabilities (0, 50, and 100, respectively). Using linear probability models, we relate these binary outcomes to the EA score. For all three events, we observe the same pattern of association: the EA score is negatively associated with providing a subjective probability indicating complete certainty (0 or 100), and is largely uncorrelated with providing a focal probability of 50 percent. The magnitudes of these associations are substantial. For example, Column [2] of Panel B suggests that a one standard deviation increase in the EA score is associated with a 0.4 percentage point reduction in the probability of reporting a 0% probability that the economy will suffer a major depression in the next 10 years. For comparison, according to Panel E in Table 8, 5% of individuals a 0% probability for this event. While we find no statistically significant association between the EA score and reporting a 100% probability that the stock market will increase, we do find a relationship between 100% beliefs about economic recessions and double-digit inflation.

[57] In results available from the authors, we show that main results relating the EA score to deviations from objective probabilities remain qualitatively similar for reasonably large intervals around the objective probabilities we use.

These results suggest that individuals with lower genetic scores report beliefs that are at odds with objective probabilities and, moreover, tend to heap on "extreme" beliefs. It is possible, however, that these reported beliefs are not related to individual behavior in a meaningful way, making these results interesting but not particularly useful for understanding the potential underlying mechanisms linking the EA score to financial decisions. This would be the case if either the HRS expectations questions do a poor job of eliciting true beliefs about these economic events, or if the events themselves were not relevant for the household's choice problem. In Appendix B, we show that these elicited beliefs do indeed predict relevant behaviors such as stock market participation, and are associated with wealth. Further, excessive optimism about the stock market is actually associated with *greater* wealth, likely due to an increase in participation. This suggests that the direction of incorrect beliefs is important for their overall impact on wealth.

We next investigate whether the EA score is associated with the length of the financial planning horizon. The complexity of economic decisions increases with their scope, and households may be heterogeneous in the costliness of thinking about increasingly distant future periods. Those for whom such considerations are relatively low-cost will endogenously consider longer horizons. The HRS asks respondents about their planning horizons for spending and saving: "In deciding how much of their (family) income to spend or save, people are likely to think about different financial planning periods. In planning your (family's) saving and spending, which of the following time periods is most important to you (and your husband/wife/partner): the next few months, the next year, the next few years, the next 5-10 years, or longer than 10 years?" Summary statistics for planning horizons are found in Panel E of Table 8 for mutually exclusive categories. 13% of respondents report planning horizons of less than 1 year. 29% have a planning horizon of a few years, 34% indicate horizons in the range of 5-10 years, and 11% have planning horizons of more than 10 years.

In Panel B of Table 13, we test whether the EA score predicts planning horizon responses. The dependent variable in Column [1] is a dummy variable equal to one if the planning horizon is greater than a few months. The estimated coefficient is statistically significant, and suggests that a one standard deviation in the EA score is associated with a 0.8 percentage point increase in the probability of reporting a planning horizon longer than a few months. Columns [2]-[4] repeat this exercise, but with dummies equal to one for increasingly longer horizons. The dummy dependent variable in Column [2] is equal to one if the reported horizon is "a few years", in [3] if "5-10 years", and in [4] if "longer than 10 years". In all but Column [4], the coefficient on the EA score is positive and significant at the 1% level. This suggests that the EA score is predictive of longer planning horizons for all but the longest horizon.

The results linking the length of the planning horizon to the EA score are consistent with an interpretation that higher EA score households are better able to think about complex and abstract decision problems. Alternatively, these results could be interpreted as an association between the EA score and patience. However, in results available from the authors and using the HRS questions designed to elicit patience parameters, we find little variation between households that report the shortest and longest planning horizons. This suggests it is unlikely the planning horizon results are due to patience. Gabaix and Laibson (2017) provide a theoretical foundation for our interpretation. They demonstrate that infinitely patient, Bayesian households that receive noisy, unbiased signals about future events will behave *as if they are impatient.* A consequence of their model is that households that receive more precise signals will appear to behave as if they are more patient than others, even though all households are equally (infinitely) patient.

5.6 Pensions

One consequence of the apparent relationship between genetic endowments and financial decisions is that individuals with low EA scores may benefit from outsourcing certain economic choices, such as saving and investment decisions. Defined benefit pensions, which may be provided by one's employer, offer one form of outsourcing by providing an employee a guaranteed stream of income in retirement without requiring the individual to choose the contribution rate or underlying investment allocations. Defined benefit plans effectively reduce the impact of the household's financial decisions on accumulated wealth by ensuring a minimal level of resources at retirement. We investigate whether the reduced autonomy associated with defined-benefit pensions alters the relationship between genetic ability and wealth.

According to Panel F of Table 8, 56% of households have a defined-benefit pension with an average present discounted value of $45,000. One primary concern is that pension participation is not randomly assigned. As a first step, we regress an indicator for defined-benefit pension participation on the average household EA score.[58] Column [1] of Table 14 shows that after including our standard set of controls, there is no economically or statistically significant relationship between the EA score and defined-benefit pension participation. Column [2] shows that, conditional on participation in a defined-benefit pension plan, defined-benefit pension wealth (the present value of pension income) is also unrelated to the EA score.[59] In general, selection into careers based on defined-benefit pension benefits

[58]Because we focus only on retired households, our definition of defined-benefit plan participation is whether the household reports receiving income from a defined-benefit pension in that household-year.

[59]We winsorize defined-benefit pension wealth at the 1^{st}, and 99^{th} percentiles.

appears to be uncorrelated with the EA score after controlling for education.

Columns [3]-[4] in Table 14 investigate whether participation in a defined-benefit plan mitigates the role of the EA score in wealth accumulation. Column [3] shows that participation in a defined-benefit plan is associated with a 39 percent increase in wealth. This is unsurprising given the inadequacy of retirement savings for many households. In Column [4], we also include an interaction between the EA score and the pension-participation dummy.[60] The results are striking. The coefficient on the interaction is negative and statistically significant, and is economically large. For households that participate in a defined-benefit plan, the coefficient on the EA score is 0.038, compared to 0.207 for households that do not participate in a defined-benefit plan. Put differently, the relationship between the EA score and wealth is over five times as large for households that have more autonomy over their savings and investment choices. This offers strong evidence in support of the hypothesis that the gene-wealth association documented in this paper is in part determined by a household's difficulty in making wise financial choices.

6 Conclusions

We study the genetic endowments linked to educational attainment, summarized as a linear index called a polygenic score (EA score). Using data from the HRS, we demonstrate that the average EA score in a household strongly and robustly predicts wealth at retirement. The estimated gene-wealth gradient is not fully explained by flexibly controlling for the education and income variables available for HRS respondents. We explore a host of additional mechanisms, including transfers from parents, risk preferences, and savings. We provide evidence that some of these factors, such as risk preferences, are related to the EA score. We fail to find evidence of a link to others, such as savings. However, we do find a strong relationship between the EA score and portfolio choices, with higher EA scores being associated with greater probabilities of business and stock ownership. Lower EA scores predict beliefs that are at odds with objective information and probabilistic thinking, as well as shorter planning horizons. Finally, the EA score is much more strongly related to wealth within a subsample of individuals — those without defined-benefit pensions — who presumably have greater autonomy over their financial decisions. Together, the associations we report provide preliminary evidence that endowments related to human capital accumulation are associated with financial wealth in part through a facility with complex dynamic decision-making.

Economic research using information on genetic endowments is useful for understanding

[60] We also include interactions between the pension-participation dummy and all principal components variables to account for possible population stratification in obtaining defined-benefit pensions.

what has heretofore been a form of unobserved heterogeneity that persists across generations, since parents provide genetic material for their children. Studies that ignore this type of heterogeneity when studying the intergenerational persistence of economic outcomes, such as income or wealth, could place too much weight on other mechanisms such as attained education or direct monetary transfers between parents and children. The use of observed genetic information helps economists to develop a more accurate and complete understanding of inequality across generations.

Moreover, studying how genetic endowments implicated in one outcome, in this case education, relate to other outcomes, such as wealth, leads to a more complete picture of how these endowments function, including how they relate to policy-relevant environmental factors (often known as "G by E" interactions). Our results on pensions and the gene-wealth gradient are an illustration of how environmental factors can modify the relationship between genetic endowments and key economic outcomes. Moreover, our results suggest that the genes implicated in educational attainment function at least in part through differences in probabilistic thinking, planning horizons and decision-making under uncertainty, all of which could also help to explain why individuals with these endowments attain more years of schooling.

In short, our study illustrates how economists can benefit from results in behavioral genetics that link specific genetic endowments to economic outcomes. The results presented here also highlight some promising next steps for research in behavioral genetics by pointing to some specific behavioral mechanisms that may link genetic endowments to economic outcomes.

References

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82 (1):137–196.

Angrist, Joshua D, Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher R Walters. 2016. "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice." *Journal of Labor Economics* 34 (2):275–318.

Arrondel, Luc, Hector F Calvo Pardo, and Derya Tas. 2014. "Subjective Return Expectations, Information and Stock Market Participation: Evidence from France." Mimeo, University of Southampton.

Bach, Stefan, Andreas Thiemann, and Aline Zucco. 2015. "The Top Tail of the Wealth Distribution in Germany, France, Spain, and Greece." DIW Berlin Discussion Paper.

Beauchamp, Jonathan P., David Cesarini, Magnus Johannesson, Matthijs J. H. M. van der Loos, Philipp D. Koellinger, Patrick J.F. Groenen, James H. Fowler, J. Niels Rosenquist, A. Roy Thurik, and Nicholas A. Christakis. 2011. "Molecular Genetics and Economics." *Journal of Economic Perspectives* 25 (4):57–82.

Becker, G.S. 1964. *Human Capital: A Theoretical Analysis with Special Reference to Education*. Columbia University Press.

Behrman, Jere R and Nancy Birdsall. 1983. "The Quality of Schooling: Quantity Alone is Misleading." *American Economic Review* 73 (5):928–946.

Belsky, Daniel W., Terrie E. Moffitt, David L. Corcoran, Benjamin Domingue, HonaLee Harrington, Sean Hogan, Renate Houts, Sandhya Ramrakha, Karen Sugden, Benjamin S. Williams, Richie Poulton, and Avshalom Caspi. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated with Educational Attainment Relate to Life-Course Development." *Psychological Science* 27:957–972.

Benhabib, Jess and Alberto Bisin. 2016. "Skewed Wealth Distributions: Theory and Empirics." NBER Working Paper.

Benhabib, Jess, Alberto Bisin, and Mi Luo. 2015. "Wealth Distribution and Social Mobility in the US: A Quantitative Approach." NBER Working Paper.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu. 2011. "The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents." *Econometrica* 79 (1):123–157.

Benham, Lee. 1974. "Benefits of Women's Education within Marriage." *Journal of Political Economy* 82 (2, Part 2):S57–S71.

Benjamin, Daniel J, David Cesarini, Christopher F Chabris, Edward L Glaeser, David I Laibson, Vilmundur Guðnason, Tamara B Harris, Lenore J Launer, Shaun Purcell, Albert Vernon Smith et al. 2012. "The Promises and Pitfalls of Genoeconomics." *Annual Review of Economics* 4:627–662.

Bierut, Laura Jean. 2010. "Convergence of Genetic Findings for Nicotine Dependence and Smoking Related Diseases with Chromosome 15q24-25." *Trends in Pharmacological Sciences* 31 (1):46–51.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital." *American Economic Review* 95 (1):437–449.

Branigan, Amelia R, Kenneth J McCallum, and Jeremy Freese. 2013. "Variation in the Heritability of Educational Attainment: An International Meta-Analysis." *Social Forces* :109–140.

Card, David and Alan B Krueger. 1996. "Labor Market Effects of School Quality: Theory and Evidence." NBER Working Paper.

Cesarini, David, Christopher T Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. 2009. "Genetic Variation in Preferences for Giving and Risk Taking." *Quarterly Journal of Economics* 124 (2):809–842.

Cesarini, David, Magnus Johannesson, Paul Lichtenstein, Örjan Sandewall, and Björn Wallace. 2010. "Genetic Variation in Financial Decision-Making." *The Journal of Finance* 65 (5):1725–1754.

Charles, Kerwin Kofi and Erik Hurst. 2003. "The Correlation of Wealth across Generations." *Journal of Political Economy* 111 (6):1155–1182.

Charles, Kerwin Kofi, Erik Hurst, and Alexandra Killewald. 2013. "Marital Sorting and Parental Wealth." *Demography* 50 (1):51–70.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9):2633–2679.

Cronqvist, Henrik and Stephan Siegel. 2014. "The Genetics of Investment Biases." *Journal of Financial Economics* 113:215–234.

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

De Nardi, Mariacristina. 2004. "Wealth Inequality and Intergenerational Links." *Review of Economic Studies* 71 (3):743–768.

Delavande, Adeline, Michael Perry, Robert Willis et al. 2006. "Probabilistic Thinking and Early Social Security Claiming." Michigan Retirement Research Center Working Paper.

Delavande, Adeline, Susann Rohwedder, and Robert J Willis. 2008. "Preparation for Retirement, Financial Literacy and Cognitive Resources." Michigan Retirement Research Center Research Paper.

Deming, David J, Justine S Hastings, Thomas J Kane, and Douglas O Staiger. 2014. "School Choice, School Quality, and Postsecondary Attainment." *American Economic Review* 104 (3):991–1013.

Dobbie, Will and Roland G Fryer Jr. 2011. "Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics* 3 (3):158–87.

Dominitz, Jeff and Charles F Manski. 2007. "Expected Equity Returns and Portfolio Choice: Evidence from the Health and Retirement Study." *Journal of the European Economic Association* 5 (2-3):369–379.

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

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2016. "Heterogeneity and Persistence in Returns to Wealth." NBER Working Paper.

Fagereng, Andreas, Magne Mogstad, Marte Rønning et al. 2015. "Why Do Wealthy Parents Have Wealthy Children." *Statistics Norway, Discussion Papers* 813.

Feuk, Lars, Andrew R. Carson, and Stephen W. Scherer. 2006. "Structural Variation in the Human Genome." *Nature Reviews Genetics* 7:85–97.

Gabaix, Xavier and David Laibson. 2017. "Myopia and Discounting." NBER Working Paper.

Grinblatt, Mark, Seppo Ikäheimo, Matti Keloharju, and Samuli Knüpfer. 2015. "IQ and Mutual Fund Choice." *Management Science* 62 (4):924–944.

Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa. 2011. "IQ and Stock Market Participation." *Journal of Finance* 66 (6):2121–2164.

Heckman, James J and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review* 91 (2):145–149.

Hudomiet, Peter, Gábor Kézdi, and Robert J Willis. 2011. "Stock Market Crash and Expectations of American Households." *Journal of Applied Econometrics* 26 (3):393–415.

Hurd, Michael D. 2009. “Subjective Probabilities in Household Surveys.” *Annual Review of Economics* 1:543.

Jackson, C Kirabo, Rucker C Johnson, and Claudia Persico. 2015. “The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms.” *Quarterly Journal of Economics* 131 (1):157–218.

Jones, Charles I. 2015. “Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality.” *The Journal of Economic Perspectives* 29 (1):29–46.

Kahneman, Daniel and Amos Tversky. 1972. “Subjective Probability: A Judgment of Representativeness.” In *The Concept of Probability in Psychological Experiments*. Springer, 25–48.

Kézdi, Gábor and Robert J Willis. 2003. “Who Becomes a Stockholder? Expectations, Subjective Uncertainty, and Asset Allocation.” Unpublished manuscript, University of Michigan.

———. 2009. “Stock Market Expectations and Portfolio Choice of American Households.” Unpublished manuscript, University of Michigan.

———. 2013. “Expectations, Aging and Cognitive Decline.” In *Discoveries in the Economics of Aging*. University of Chicago Press, 305–337.

Kimball, MS, CR Sahm, and MD Shapiro. 2009. “Risk Preferences in the PSID: Individual Imputations and Family Covariation.” *American Economic Review P&P* 99 (2):363–368.

Kindermann, Fabian and Dirk Krueger. 2014. “High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk.” NBER Working Paper.

Lee, J. J. and et al. 2018. “Gene Discovery and Polygenic Prediction from a 1.1-Million-Person GWAS of Educational Attainment.” *Nature Genetics*, forthcoming.

Lillard, Lee and Robert J Willis. 2001. “Cognition and Wealth: The Importance of Probabilistic Thinking.” *Michigan Retirement Research Center Working Paper* .

Locke, Adam E., Bratati Kahali, Sonja I. Berndt, Anne E. Justice, Tune H. Pers et al. 2015. “Genetic Studies of Body Mass Index Yield New Insights for Obesity Biology.” *Nature* 518 (7538):197–206.

Lumsdaine, Robin L and Rogier J D Potter van Loon. 2017. “Do Survey Probabilities Match Financial Market Beliefs?” *Journal of Behavioral Finance*, forthcoming.

Lusardi, Annamaria, Pierre-Carl Michaud, and Olivia S Mitchell. 2017. "Optimal Financial Knowledge and Wealth Inequality." *Journal of Political Economy*, forthcoming.

Martin, Alicia R, Christopher R Gignoux, Raymond K Walters, Genevieve L Wojcik, Benjamin M Neale, Simon Gravel, Mark J Daly, Carlos D Bustamante, and Eimear E Kenny. 2017. "Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations." *The American Journal of Human Genetics* 100 (4):635–649.

Okbay, Aysu, Jonathan P Beauchamp, Mark Alan Fontana, James J Lee, Tune H Pers, Cornelius A Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson, S Fleur W Meddens et al. 2016. "Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment." *Nature* 533 (7604):539–542.

Okbay, Aysu, Daniel Benjamin, and Peter Visscher. 2018. "SSGAC Educational Attainment: GWAS and MTAG Polygenic Scores (Ver. 1.0)." SSGAC Educational Attainment: GWAS and MTAG Polygenic Scores (Ver. 1.0).

Olsen, Anya and Russell Hudson. 2009. "Social Security Administration's Master Earnings File: Background Information." *Social Security Bulletin* 69 (3). URL `https://www.ssa.gov/policy/docs/ssb/v69n3/v69n3p29.html`.

Papageorge, Nicholas W and Kevin Thom. 2016. "Genes, Education and Labor Outcomes: Evidence from the Health and Retirement Study." IZA Discussion Paper 10200.

Poterba, James M and David A Wise. 1998. "Individual Financial Decisions in Retirement Saving Plans and the Provision of Resources for Retirement." In *Privatizing Social Security*. University of Chicago Press, 363–401.

Price, Alkes L, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nature Genetics* 38 (8):904–909.

Rietveld, Cornelius A, Sarah E Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W Martin, Harm-Jan Westra, Konstantin Shakhbazov, Abdel Abdellaoui, Arpana Agrawal et al. 2013. "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment." *Science* 340 (6139):1467–1471.

Rohwedder, Susann and Robert J Willis. 2010. "Mental Retirement." *The Journal of Economic Perspectives* 24 (1):119–138.

Saez, Emmanuel and Gabriel Zucman. 2014. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data."

Savage, L.J. 1954. *The Foundations of Statistics*. Wiley.

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

Smith-Woolley, Emily, Jean-Baptiste Pingault, Saskia Selzam, Kaili Rimfeld, Eva Krapohl, Sophie von Stumm, Kathryn Asbury, Philip S Dale, Toby Young, Rebecca Allen et al. 2018. "Differences in Exam Performance between Pupils Attending Selective and Non-Selective Schools Mirror the Genetic Differences between Them." *npj Science of Learning* 3 (1):3.

Thorgeirsson, Thorgeir E, Daniel F Gudbjartsson, Ida Surakka, Jacqueline M Vink, Najaf Amin, Frank Geller, Patrick Sulem, Thorunn Rafnar, Tõnu Esko, Stefan Walter et al. 2010. "Sequence Variants at CHRNB3-CHRNA6 and CYP2A6 Affect Smoking Behavior." *Nature Genetics* 42 (5):448–453.

Tucker-Drob, Elliot M. and Timothy C. Bates. 2015. "Large Cross-National Differences in Gene x Socioeconomic Status Interaction on Intelligence." *Psychological Science* 27 (2):138–149.

Venti, Steven F and David A Wise. 1998. "The Cause of Wealth Dispersion at Retirement: Choice or Chance?" *American Economic Review* 88 (2):185–191.

Vilhjalmsson, et al., Bjarni J. 2015. "Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores." *The American Journal of Human Genetics* 87:576–592.

Wiswall, Matthew and Basit Zafar. 2015. "Determinants of College Major Choice: Identification Using an Information Experiment." *Review of Economic Studies* 82 (2):791–824.

Yitzhaki, Shlomo. 1987. "The Relation between Return and Income." *Quarterly Journal of Economics* 102 (1):77–95.

Yogo, Motohiro. 2016. "Portfolio Choice in Retirement: Health Risk and the Demand for Annuities, Housing, and Risky Assets." *Journal of Monetary Economics* 80:17–34.

7 Tables and Figures

Table 1: Summary Statistics — Birth Year and Education

	All Households (4297)		Coupled (3026)		Female Only (309)		Male Only (962)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Year of birth								
Female	1934.41	9.47	1935.32	8.86	1931.53	10.69	.	.
Male	1932.39	8.91	1932.07	8.68	.	.	1935.49	10.40
Years of Education								
Female	12.68	2.35	12.73	2.32	12.50	2.45	.	.
Male	12.77	2.95	12.77	2.95	.	.	12.78	3.01
No Degree								
Female	0.17	0.37	0.16	0.36	0.20	0.40	.	.
Male	0.19	0.40	0.20	0.40	.	.	0.16	0.37
GED								
Female	0.04	0.20	0.04	0.20	0.04	0.19	.	.
Male	0.06	0.23	0.05	0.23	.	.	0.07	0.26
High School								
Female	0.58	0.49	0.59	0.49	0.57	0.49	.	.
Male	0.48	0.50	0.47	0.50	.	.	0.49	0.50
2-Year Degree								
Female	0.04	0.20	0.04	0.20	0.04	0.20	.	.
Male	0.03	0.17	0.03	0.17	.	.	0.05	0.21
4-Year Degree								
Female	0.11	0.31	0.11	0.32	0.08	0.27	.	.
Male	0.13	0.34	0.13	0.34	.	.	0.14	0.34
Masters								
Female	0.05	0.22	0.05	0.22	0.06	0.23	.	.
Male	0.08	0.27	0.08	0.27	.	.	0.07	0.25
Professional								
Female	0.01	0.08	0.00	0.06	0.01	0.11	.	.
Male	0.03	0.17	0.03	0.18	.	.	0.02	0.15
Some Post-Secondary / Unknown								
Female	0.00	0.03	0.00	0.03	0.00	0.00	.	.
Male	0.00	0.03	0.00	0.03	.	.	0.01	0.08

Notes: This table reports means and standard deviations (SD) for birth year, years of education, and indicator variables for highest degree obtained for s household members examined in our main analytic sample. The final category "Some Post Secondary / Unknown" includes some education after high school, but no completed or recorded degree, and cases with unknown degree. Summary statistics are presented for males and females in all households and then separately for those in coupled households (where both members are observed at least once), female-only households, and male-only households.

Table 2: Summary Statistics — Income

Panel A	All Households		Coupled		Female Only		Male Only	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Non-Missing Income	0.90	0.30	0.90	0.30	0.89	0.31	0.90	0.31
Zero Income	0.01	0.07	0.00	0.03	0.02	0.13	0.01	0.08
Avg. Yrs. Top-Coded	12.19	12.57	15.06	12.55	3.50	7.59	10.78	12.56
Never Top Coded	0.29	0.46	0.17	0.38	0.69	0.46	0.26	0.44
Tot. Real Inc.	2171.12	1405.13	2476.31	1376.18	1301.01	1082.62	1863.10	1380.59
	Quantiles of Income							
Panel B	10	25	50	75	90	Mean	SD	N
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
All Households	429.36	1122.23	2065.58	2963.52	3941.04	2171.12	1405.13	3853
Coupled	773.76	1532.96	2391.92	3208.35	4191.20	2476.31	1376.18	2720
Female Only	134.86	441.01	1084.00	1874.06	2704.72	1301.01	1082.62	856
Male Only	396.46	938.94	1660.59	2494.59	3250.70	1863.10	1380.59	277

Notes: This table reports summary statistics for the income measure used in our main analysis obtained from the Social Security Administration Master Earnings File data. Panel A provides information on the proportion of households with non-missing data, the proportion of households with zero income (conditional on having non-missing income data), average number of years that any household member reports top-coded income, the fraction of households never observed with a top-coded income observation, and average total real income (in $1,000s). In years in which a household member's income is top-coded, we replace the top-coded amount with the average of individual earned incomes greater than or equal to the top-coded amount in the Current Population Survey for that year. Income statistics are provided for all households and separately by household structure. Panel B reports quantiles of the income distribution (in $1000s) along with the mean and standard deviation of income by household structure.

Table 3: Summary Statistics — Wealth

	p10	p25	p50	p75	p90	Mean	SD
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Total Wealth	40.73	124.72	321.04	687.53	1311.97	563.06	733.98
No Housing	17.66	55.45	179.57	474.16	1000.68	459.95	1430.83
No Ret.	5.41	77.34	234.33	562.41	1142.00	518.07	1203.57
No H or R	0.00	8.00	90.86	339.28	825.98	348.00	1029.78

Notes: This table reports summary statistics on the distribution of real wealth in our main sample. Measures of wealth include total real wealth, total wealth excluding housing, total wealth excluding retirement wealth (the present discounted value of retirement income), and the total wealth excluding housing and retirement wealth. Each wealth measure is winsorized at the 1st and 99th percentiles. These statistics are calculated for the full analytical sample of 15,670 household-year observations.

Table 4: EA Score Related to Key Economic Variables

Panel A
Individual EA Score
and Individual Variables

	Q1	Q2	Q3	Q4	Q4−Q1 p-value
	[1]	[2]	[3]	[4]	[5]
Birth Year	1935.13	1934.64	1933.53	1932.81	0.00
Male	0.38	0.39	0.40	0.42	0.02
Father's Education	8.95	9.55	9.95	10.58	0.00
Mother's Education	9.49	10.16	10.42	10.93	0.00
Education	11.89	12.52	13.11	13.88	0.00
Age	72.11	72.87	73.78	74.15	0.00
Coupled	0.45	0.47	0.52	0.55	0.00

Panel B
Avg. Household EA Score
and Individual Variables

	Q1	Q2	Q3	Q4	Q4−Q1 p-value
	[1]	[2]	[3]	[4]	[5]
Female:					
Birth Year	1935.50	1934.87	1934.38	1932.86	0.00
Education	11.78	12.29	13.01	13.62	0.00
Father's Education	8.75	9.25	10.02	10.70	0.00
Mother's Education	9.34	9.84	10.50	10.82	0.00
Age	70.81	71.32	71.92	73.06	0.00
Male:					
Birth Year	1933.70	1932.77	1932.24	1930.88	0.00
Education	11.65	12.45	13.08	13.87	0.00
Father's Education	8.73	9.30	9.84	10.51	0.00
Mother's Education	9.30	9.87	10.33	10.83	0.00
Age	71.35	72.49	73.19	74.01	0.00
Household:					
Avg. HH Income (in $1000)	1989.00	2135.81	2263.04	2297.20	0.00

Notes: This table relates the EA score to key economic variables. Columns [1]-[4] separate individuals into quartiles of the individual EA score distribution and report average values of demographic variables for each quartile. Averages are calculated for the 5,692 genotyped individuals belonging to a household in the sample; age and coupled status are time-varying and thus their averages are calculated over the larger person-year observation count. Column [5] reports the p-value associated with test of difference in means between the fourth and first quartiles. Panel B repeats this exercise at the household level and separates households into quartiles of the average household EA score. We report averages for individual-level variables, separately for females and males, for the full analytic sample of 4,297 households. The age of male and female household members is time-varying, so averages of this variable are based on a larger sample of household-year observations. We only use age observations on living household members for these averages. The final row reports averages of our total household income variable.

Table 5: EA Score and Assortative Mating

	Panel A: Unadjusted means (N=1,477)				**Panel B:** Adjusted for degree and years of education (N=1,473)			
	Female EA Quartile				Female EA Quartile			
Male EA Quartile	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
Q1	29.2	27.0	21.8	20.3	29.3	20.2	24.6	23.2
Q2	28.2	22.2	25.3	23.1	25.3	27.8	26.8	23.7
Q3	24.7	26.7	25.1	23.7	24.5	24.7	25.1	26.1
Q4	18.0	24.1	27.8	32.9	21.0	27.2	23.5	27.0

Notes: This table reports the distribution of the male household member's EA score conditional on the quartile of the female household member's EA score for all coupled households with non-missing EA scores for both members. For each panel, each row-column entry reports the probability that a female with an EA score in the quartile corresponding to the column is coupled with a male whose EA score is in the quartile corresponding to the row. The column probabilities sum to 100 percent. Panel A presents these statistics based unconditional individual EA scores. Panel B presents the same statistics based on the residual EA score obtained from a regression of the individual EA score on years of education and indicators for highest degree attained.

Table 6: Average Household EA Score and Household Wealth

Dep. Var: Log Wealth	[1]	[2]	[3]	[4]	[5]	[6]	[7]
EA Score	0.331***	0.319***	0.305***	0.141***	0.124***	0.291***	0.115***
	(0.023)	(0.021)	(0.021)	(0.022)	(0.021)	(0.022)	(0.022)
Male Educ				0.091***			
				(0.009)			
Female Educ				0.148***			
				(0.010)			
Log Income						0.263***	0.215***
						(0.032)	(0.027)
Obs.	15670	15670	15670	15660	15660	14273	14266
R^2	0.048	0.250	0.270	0.352	0.406	0.306	0.434
Standard Controls		X	X	X	X	X	X
Principal Comp.			X	X	X	X	X
Years of Educ.				X			
Full Educ. Controls					X		X

Notes: This table presents estimates from regressions of log household wealth on average household EA score and varying sets of controls. Column [1] includes no controls. Column [2] includes controls for age, birth cohort, sex of respondent, and calendar year, as described in Section 4.2. Column [3] adds controls for principal components of the genetic data for genotyped household members. Column [4] adds years of education separately for both female and male household members. Column [5] replaces the two schooling variables with our full set of education controls (dummies for years of education, degree dummies and interactions as described in Section 4.2). Column [6] includes the log of total household income, but excludes any controls for education. Column [7] includes our full set of controls including the detailed education variables and the log of total household income. Significance stars ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered at the family level.

Table 7: EA Score and Household Wealth: Alternative Specifications

Panel A **FR and Non-FR**	[1]	[2]	[3]	[4]	[5]	[6]	[7]
FR EA Score	0.255***	0.272***	0.268***	0.139***	0.128***	0.252***	0.116***
	(0.028)	(0.028)	(0.028)	(0.026)	(0.028)	(0.028)	(0.029)
Non-FR EA Score	0.156***	0.160***	0.159***	0.052*	0.046*	0.152***	0.047*
	(0.030)	(0.028)	(0.028)	(0.027)	(0.026)	(0.028)	(0.026)
Male Educ.				0.091***			
				(0.014)			
Female Educ.				0.109***			
				(0.015)			
Log Income						0.232***	0.190***
						(0.055)	(0.048)
Obs.	5350	5350	5350	5340	5340	5065	5058
R^2	0.071	0.216	0.225	0.331	0.448	0.245	0.462
Panel B **Include Non-Retired**	[1]	[2]	[3]	[4]	[5]	[6]	[7]
EA Score	0.319***	0.311***	0.305***	0.139***	0.116***	0.275***	0.104***
	(0.018)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)
Male Educ.				0.093***			
				(0.008)			
Female Educ.				0.135***			
				(0.008)			
Log Income						0.299***	0.238***
						(0.028)	(0.024)
Obs.	27845	27845	27845	27795	27795	25423	25384
R^2	0.044	0.209	0.222	0.298	0.339	0.259	0.363
Panel C: **SSA and HRS Income**	[1]	[2]	[3]	[4]	[5]	[6]	[7]
EA Score	0.344***	0.329***	0.329***	0.152***	0.122***	0.266***	0.108***
	(0.029)	(0.027)	(0.027)	(0.028)	(0.028)	(0.027)	(0.028)
Male Educ.				0.088***			
				(0.012)			
Female Educ.				0.134***			
				(0.013)			
Log Income (SSA)						0.223***	0.220***
						(0.051)	(0.041)
Log Income (HRS)						0.375***	0.242***
						(0.040)	(0.034)
Obs.	7998	7998	7998	7994	7994	7588	7587
R^2	0.059	0.265	0.300	0.384	0.452	0.376	0.491
Standard Controls		X	X	X	X	X	X
Principal Comp.			X	X	X	X	X
Years of Educ.				X			
Full Educ. Controls					X		X
Log Income						X	X

Notes: This table presents estimates from regressions of log household wealth on household polygenic scores and varying sets of controls. (see notes to Table 6 for an explanation of the controls in each column). In Panel A, we include the individual EA score of the financial respondent (FR) and the score of the non-financial respondent (NFR) separately and restrict the sample to households with two genotyped members. Panel B uses the average EA score, but expands our sample to include households that are not yet retired. In Panel C, we again restrict our attention to retired households, but now include two separate household incomes measures: our primary measure derived from the Social Security Administration data, and the log of average annual income based on self-reported values in the HRS. In Panel C, all specifications are restricted to households with non-missing observations of the HRS income variable. Significance stars ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered at the family level.

Table 8: Summary Statistics: Mechanisms

	All Households			Coupled			Female Only			Male Only		
	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
Panel A: Transfers	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Any Inheritance	0.34	0.47	4297	0.38	0.49	3026	0.28	0.45	309	0.23	0.42	962
Inheritance Amount (in $1000)	143.20	528.64	8717	144.76	559.09	6615	172.76	614.08	521	123.25	255.61	1581
Fathers' Education (HH Avg.)	9.61	3.23	3271	9.52	3.05	2498	10.20	3.79	222	9.75	3.68	551
Mothers' Educ (HH Avg.)	10.09	2.81	3362	10.02	2.65	2516	10.65	3.12	235	10.20	3.24	611
Panel B: Mortality and Savings	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Mortality	0.03	0.18	33297	0.03	0.18	27131	0.05	0.23	1432	0.04	0.20	4734
Mortality Expectations	67.39	26.03	29430	67.78	25.66	25123	59.83	28.97	1182	67.13	27.40	3125
Expenditures (in $1000)	56.36	38.90	3186	60.81	39.99	2585	38.11	29.73	160	36.95	25.02	441
Panel C: Risk Aversion (Income)	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Not take 50-50 Gamble												
Doubling Income or 10% Cut	0.39	0.49	10632	0.39	0.49	9138	0.38	0.49	425	0.40	0.49	1069
Take 50-50 Gamble												
Doubling Income or												
10% Cut (but not 20%)	0.22	0.42	10632	0.22	0.42	9138	0.19	0.40	425	0.22	0.41	1069
20% Cut (but not 33%)	0.17	0.37	10632	0.17	0.38	9138	0.13	0.34	425	0.16	0.37	1069
33% Cut (but not 50%)	0.10	0.29	10632	0.10	0.29	9138	0.11	0.32	425	0.09	0.29	1069
50% Cut (but not 75%)	0.07	0.26	10632	0.07	0.26	9138	0.10	0.30	425	0.06	0.25	1069
75% Cut	0.05	0.22	10632	0.05	0.21	9138	0.09	0.29	425	0.06	0.23	1069

Notes: This table reports means and standard deviations for additional variables used to investigate mechanisms underlying the estimated gene-wealth gradient. Each panel corresponds to an alternative mechanism. Mechanisms include Transfers (Panel A); Savings (Panel B); Risk aversion (Panel C); Portfolio choices (Panel D); Beliefs and planning horizons (Panel E); and Defined-benefit pensions (Panel F). Summary statistics are reported for all households and then separately for coupled households (where a male and a female are present and genotyped), female only and male only households.

Table 8: Summary Statistics: Mechanisms (continued)

	All Households			Coupled			Female Only			Male Only		
	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
Panel D: Portfolio Choices	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Has House	0.82	0.39	17721	0.89	0.32	12705	0.67	0.47	1163	0.64	0.48	3853
Has Business	0.08	0.27	17721	0.09	0.29	12705	0.08	0.28	1163	0.05	0.21	3853
Any Stocks	0.44	0.50	14637	0.50	0.50	10274	0.37	0.48	1022	0.29	0.45	3341
Panel E: Beliefs and Planning Horizons	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Prob: Stock Market Up												
Reported Probability	48.24	26.14	36189	48.93	26.08	30605	48.55	27.01	1474	42.99	25.63	4110
Deviation from Objective	30.02	21.11	36189	29.51	20.93	30605	30.17	21.60	1474	33.76	21.88	4110
Report 0%	0.05	0.22	36189	0.05	0.21	30605	0.06	0.24	1474	0.08	0.27	4110
Report 50%	0.30	0.46	36189	0.30	0.46	30605	0.30	0.46	1474	0.31	0.46	4110
Report 100%	0.04	0.20	36189	0.04	0.20	30605	0.05	0.21	1474	0.02	0.15	4110
Planning Horizon:												
Less than 1 Year	0.13	0.34	28032	0.12	0.32	23964	0.20	0.40	989	0.21	0.41	3079
More than 1 Year	0.12	0.33	28032	0.12	0.32	23964	0.14	0.35	989	0.15	0.36	3079
More than a Few Years	0.29	0.46	28032	0.30	0.46	23964	0.25	0.43	989	0.26	0.44	3079
5-10 Years	0.34	0.47	28032	0.35	0.48	23964	0.32	0.47	989	0.28	0.45	3079
More than 10 Years	0.11	0.32	28032	0.12	0.32	23964	0.09	0.29	989	0.09	0.29	3079
Panel F: Pensions	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Has DB Pension	0.56	0.50	15660	0.60	0.49	11087	0.48	0.50	1090	0.45	0.50	3483
Pension Value (in $1000)	45.43	64.67	8717	49.95	67.84	6615	34.29	49.41	521	30.20	51.41	1581

Notes: This table reports means and standard deviations for additional variables used to investigate mechanisms underlying the estimated gene-wealth gradient. Each panel corresponds to an alternative mechanism. Mechanisms include Transfers (Panel A); Savings (Panel B); Risk aversion (Panel C); Portfolio choices (Panel D); Beliefs and planning horizons (Panel E); and Defined-benefit pensions (Panel F). Summary statistics are reported for all households and then separately for coupled households (where a male and a female are present and genotyped), female only and male only households.

Table 9: Transfers: Inheritances and Parental Education

Dep. Var:	**Receive Inheritance**	**Inheritance Amount**	**Fathers' Educ.**	**Mothers' Educ.**	**Fathers' Educ.**	**Mothers' Educ.**
	[1]	[2]	[3]	[4]	[5]	[6]
EA Score	0.021***	0.070	0.836***	0.658***	0.328***	0.193***
	(0.008)	(0.060)	(0.061)	(0.052)	(0.063)	(0.054)
Obs.	4293	1463	3271	3362	3267	3358
R^2	0.267	0.427	0.211	0.214	0.406	0.419
Standard Controls	X	X	X	X	X	X
Principal Comp.	X	X	X	X	X	X
Full Educ. Controls	X	X			X	X

Notes: This table presents estimates from regressions of inheritance and parental education variables on the average household EA score and various controls. In Column [1], the dependent variable is an binary that takes a value of 1 if the household ever receives an inheritance in the sample. In Column [2], the dependent variable is the log of the total real inheritance amount that the household receives over the course of the sample, conditional on having received an inheritance. In Column [3], the dependent variable is average years of fathers' education (averaging over household members). Column [4] repeats this exercise for average years of mothers' education. Columns [5] and [6] repeat the analysis in Columns [3] and [4], but now include controls for education of household members. Significance stars ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered at the family level.

Table 10: Mortality and Savings

Panel A: Dep. Var:	Observed Mortality			Exp. Mortality Pr(Live to 75)		
	All Ind.	**Females**	**Males**	**All Ind.**	**Females**	**Males**
	[1]	[2]	[3]	[4]	[5]	[6]
Ind. EA Score	-0.003**	-0.003**	-0.001	0.444	0.658*	-0.278
	(0.001)	(0.001)	(0.002)	(0.284)	(0.368)	(0.479)
Obs.	33297	20415	12882	29430	17615	11815
R^2	0.024	0.027	0.034	0.117	0.128	0.148
Standard Controls	X	X	X	X	X	X
Principal Comp.	X	X	X	X	X	X
Full Educ. Controls	X	X	X	X	X	X
Panel B: Dep. Var: Log HH Expenditures	**All**	**Age <60**	**Drop Low Income**	**All**	**Age <60**	**Drop Low Income**
	[1]	[2]	[3]	[4]	[5]	[6]
EA Score	0.009	0.032	0.041	-0.005	0.022	0.014
	(0.017)	(0.028)	(0.028)	(0.017)	(0.028)	(0.028)
Log Income (2nd Quintile)	-0.045	-0.082	0.000	-0.040	-0.096	0.000
	(0.038)	(0.101)		(0.040)	(0.113)	
Log Income (3rd Quintile)	0.072*	-0.048	0.030	0.094**	-0.012	0.073
	(0.040)	(0.093)	(0.076)	(0.041)	(0.095)	(0.083)
Log Income (4th Quintile)	0.118***	0.050	0.119	0.125***	0.016	0.093
	(0.042)	(0.089)	(0.075)	(0.043)	(0.092)	(0.085)
Log Income (5th Quintile)	0.227***	0.236***	0.290***	0.191***	0.189**	0.252***
	(0.049)	(0.090)	(0.077)	(0.049)	(0.090)	(0.083)
Log Wealth				0.112***	0.095***	0.063***
				(0.012)	(0.018)	(0.020)
Obs.	3186	1130	1021	2654	925	840
R^2	0.483	0.609	0.637	0.532	0.653	0.681
Standard Controls	X	X	X	X	X	X
Principal Comp.	X	X	X	X	X	X
Full Educ. Controls	X	X	X	X	X	X
Log Wealth					X	X

Notes: This table presents estimates from regressions of observed mortality, mortality expectations, and the log of household expenditures on the EA score and various controls. In Panel A, we estimate linear probability models for the annual probability of death as a function of the individual EA score. Column [1] presents estimates pooling all individuals and Columns [2] and [3] repeat this exercise separately for females and males, respectively. In Columns [4]-[6], the outcome variable is an individual's subjective probability of surviving until age 75, and again we estimate this specification for the full sample and then separately for females and males. In Panel [B], the outcome variable is the log of household expenditures, and we restrict our sample to non-retired households. In Column [1] we control for quintiles in contemporaneous annual household income. In Column [2], we restrict the sample to households with a maximum age of less than 60. In Column [3], we restrict to households with a maximum member age of less than 60, and also drop households in the lowest quintile of contemporaneous income. Columns [4]-[6] repeat the analyses in Columns [1]-[3], controlling for log wealth in the previous HRS survey year. Significance stars ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered at the family level.

Table 11: Risk Aversion

Dep. Var:	**Risk Aversion: Indicator**			**Risk Aversion: Categorical**		
	Income	**Inheritance**	**Business**	**Income**	**Inheritance**	**Business**
	[1]	[2]	[3]	[4]	[5]	[6]
EA Score	-0.022***	-0.002	-0.026**	-0.045***	0.021	-0.056**
	(0.007)	(0.012)	(0.011)	(0.015)	(0.028)	(0.027)
Obs.	10632	2972	2937	10632	2972	2937
R^2	0.107	0.210	0.246			
Standard Controls	X	X	X	X	X	X
Principal Comp.	X	X	X	X	X	X
Full Educ. Controls	X	X	X	X	X	X
Log Income	X	X	X	X	X	X

Notes: This table presents estimates from regressions of measures of individual risk tolerance on the EA score and various controls. Risk tolerance is elicited from questions based on risky gambles over labor income, inheritance wealth, and business wealth. In Columns [1]-[3], the dependent variable is an indicator that takes a value of 1 for individuals that never choose the risky option over a guaranteed outcome. In Columns [4]-[6] we report estimates from ordered probit models where the outcome is a categorical variable that takes one of six values depending on the riskiest gamble that an individual accepts, with higher values indicating greater risk aversion. Significance stars ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Table 12: Average Household EA Score and Portfolio Decisions

Panel A Dep. Var:	**Owns House** [1]	**Owns Business** [2]	**Owns Stocks** [3]	**Owns House** [4]	**Owns Business** [5]	**Owns Stocks** [6]
EA Score	0.003	0.008*	0.048***	-0.012**	0.001	0.035***
	(0.006)	(0.004)	(0.007)	(0.006)	(0.004)	(0.007)
Log Income	0.030***	-0.005	0.053***	0.005	-0.015***	0.032***
	(0.006)	(0.004)	(0.007)	(0.005)	(0.005)	(0.007)
Lagged Log Wealth				0.122***	0.043***	0.134***
				(0.004)	(0.003)	(0.006)
Obs.	17721	17721	14637	12559	12559	11278
R^2	0.240	0.111	0.291	0.370	0.165	0.399
Standard Controls	X	X	X	X	X	X
Principal Comp.	X	X	X	X	X	X
Full Educ. Controls	X	X	X	X	X	X

Panel B Dep. Var: Log Wealth	[1]	[2]	[3]	[4]	[5]	[6]
EA Score	0.124***	0.113***	0.117***	0.116***	0.072***	0.065***
	(0.023)	(0.023)	(0.019)	(0.022)	(0.021)	(0.017)
Has House			1.561***			1.364***
			(0.052)			(0.048)
Has Business				0.942***		0.715***
				(0.049)		(0.043)
Any Stocks					1.018***	0.814***
					(0.034)	(0.029)
Any Inheritance		-1.311***				-1.061***
		(0.201)				(0.155)
Log Total Inheritance		0.160***				0.121***
		(0.018)				(0.014)
Father Education (Male)		-0.005				-0.007
		(0.008)				(0.006)
Father Education (Female)		0.015*				0.002
		(0.008)				(0.006)
Mother Education (Male)		0.013				0.007
		(0.008)				(0.007)
Mother Education (Female)		-0.007				0.001
		(0.009)				(0.007)
Obs.	12927	12927	12927	12927	12927	12927
R^2	0.446	0.465	0.563	0.468	0.523	0.638
Standard Controls	X	X	X	X	X	X
Principal Comp.	X	X	X	X	X	X
Full Educ. Controls	X	X	X	X	X	X
Log Income	X	X	X	X	X	X

Notes: Panel A of this table presents estimates from regressions of indicators for ownership of different asset types on the EA score and various controls. Panel B presents estimates from regressions of log household wealth on the EA score, inheritance variables, parental education, indicators for ownership of different asset types, and various controls. Significance stars ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered at the family level.

Table 13: Extreme Beliefs and Planning Horizons

Panel A	**The EA Score and Beliefs**			
Dep. Var:	**Deviation**	**Prob=0**	**Prob=0.5**	**Prob=1**
	[1]	[2]	[3]	[4]
	Stock Market Goes Up			
EA Score	-0.572***	-0.005***	-0.003	-0.002
	(0.170)	(0.002)	(0.003)	(0.001)
Obs.	36189	36189	36189	36189
R^2	0.101	0.062	0.030	0.048
	Depression			
EA Score	-0.537***	-0.004**	-0.004	-0.008***
	(0.136)	(0.002)	(0.003)	(0.002)
Obs.	36261	36261	36261	36261
R^2	0.088	0.046	0.037	0.073
	Double Digit Inflation			
EA Score	-1.047***	-0.004**	-0.005	-0.011***
	(0.191)	(0.002)	(0.004)	(0.002)
Obs.	22786	22786	22786	22786
R^2	0.080	0.057	0.044	0.072
Panel B	**The EA Score and Planning Horizons**			
Dep. Var:	**PH$\geq$ 1 Yr.**	**PH$\geq$ Few Yrs.**	**PH$\geq$ 5-10 Yrs.**	**PH$>$ 10**
	[1]	[2]	[3]	[4]
EA Score	0.008***	0.012***	0.013***	0.004
	(0.003)	(0.004)	(0.004)	(0.003)
Obs.	28032	28032	28032	28032
R^2	0.072	0.080	0.077	0.045
Standard Controls	X	X	X	X
Principal Comp.	X	X	X	X
Full Educ. Controls	X	X	X	X

Notes: Panel A of this table presents estimates from regressions of beliefs about probabilities of three macroeconomic events on the EA score and various controls. Separate estimates are given for three distinct macroeconomic events: an increase in the stock market over the next year, a major depression in the next 10 years, and double-digit inflation in the next 10 years. In Column [1] the dependent variable is the absolute value of the deviation of the respondent's belief from an "objective" probability (as described in Section 5.5). The outcome variables in Columns [2], [3] and [4] are indicators for providing subjective probabilities of 0, 0.5 and 1, respectively. Panel B presents estimates from regressions of indicator variables for the length of a respondent's financial planning horizon on the EA score and various controls. In Column [1] the dependent variable is an indicator for reporting a planning horizon greater than or equal to one year. In Columns [2], [3] and [4], the dependent variables are indicators for horizons of "greater than or equal to a few years;" "greater than or equal to 5-10 years;" and "greater than 10 years," respectively.

Table 14: Pensions and Household Wealth

Dep. Var:	**Has Pension** [1]	**Pension Wealth** [2]	**Log Wealth** [3]	**Log Wealth** [4]
EA Score	0.012	0.004	0.120***	0.207***
	(0.008)	(0.023)	(0.021)	(0.031)
DB Pension			0.385***	0.186***
			(0.034)	(0.049)
EA Score × DB Pension				-0.169***
				(0.034)
Obs.	15660	8717	15660	15660
R^2	0.168	0.695	0.419	0.429
Standard Controls	X	X	X	X
Principal Comp.	X	X	X	X
Full Educ. Controls	X	X	X	X
Log Income				

Notes: Columns [1]-[2] of this table present estimates from regressions of defined benefit (DB) pension participation and log pension wealth (conditional on participation) on the EA score and various controls. Columns [3]-[4] presents estimates from regressions of log household wealth on the EA score, DB pension participation, an interaction between the EA score and pension participation, and various controls. Significance stars ***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.10 levels, respectively. Standard errors are clustered at the family level.

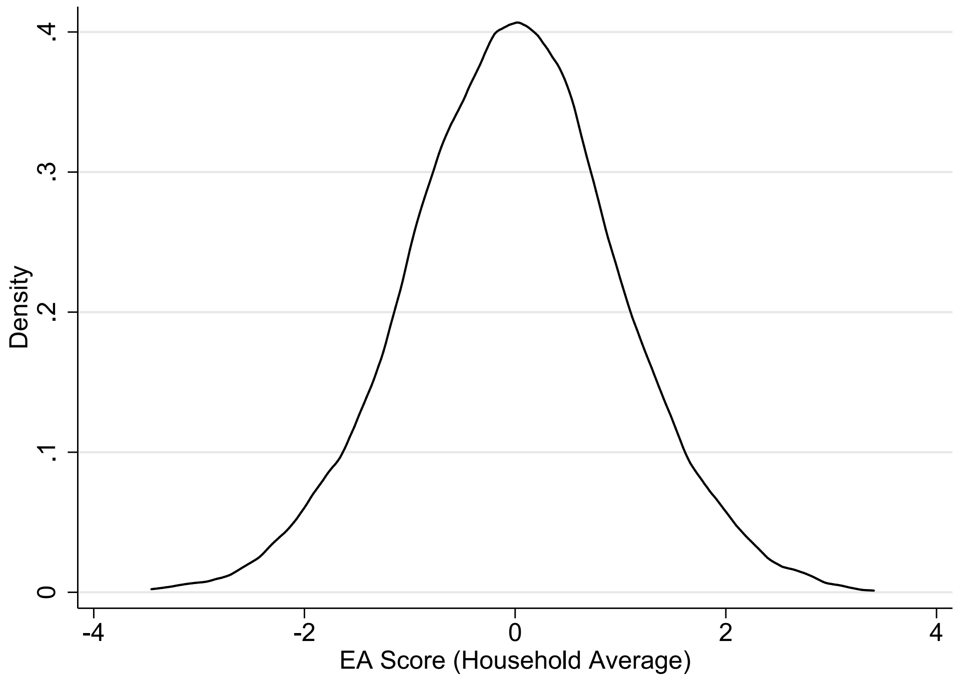

Figure 1: Distribution of Household Average EA Score

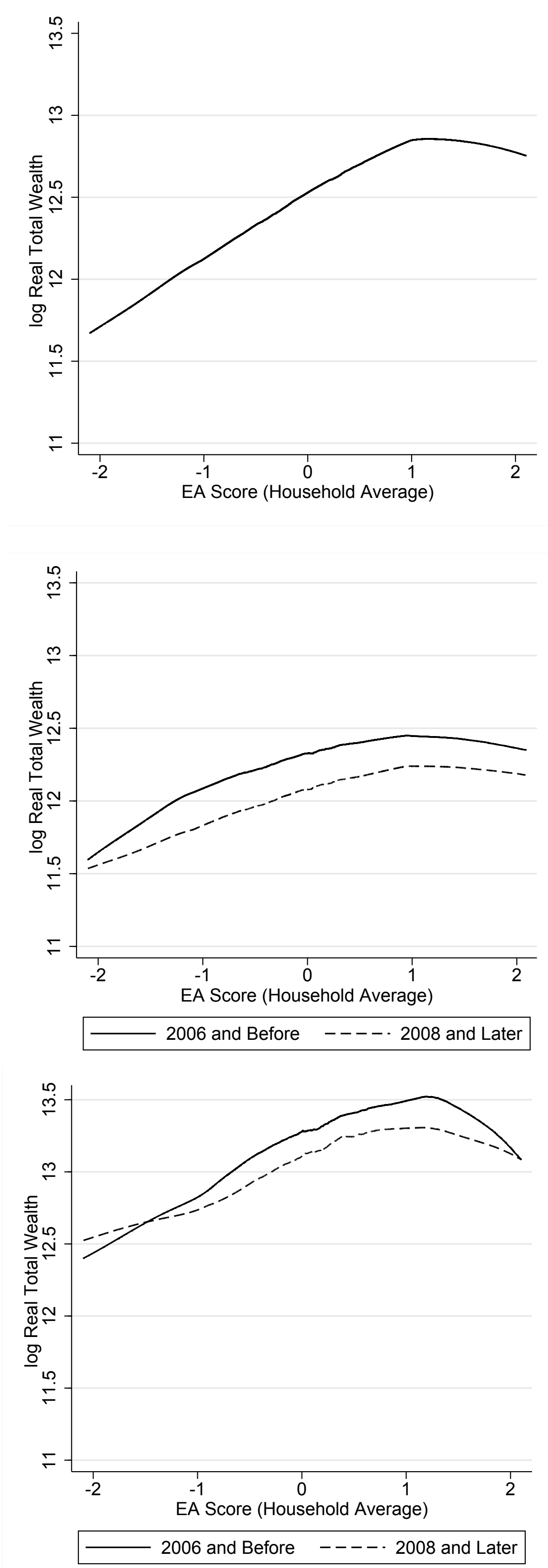

Figure 2: *Notes*: Panel A plots average household EA score v.s. log household wealth using data for all household-year observations in the analytic sample. Panel B restricts the sample to households with a maximum education level of a high school degree or less and provides separate plots for household-year observations prior to 2008, and those from 2008 or later. Panel C is analogous to Panel B, but restricts the sample to households with at least one member having at least a college degree.