The DNA of Enterprise: Genetic Endowments, Entrepreneurship and Intergenerational Mobility*

Qianyu Yang† Naijia Guo‡ Weilong Zhang§

June 26, 2026

Abstract

Are entrepreneurs born or made? This paper studies how genetic endowments shape entrepreneurship, intergenerational mobility, and wealth concentration by combining polygenic scores with a structural overlapping-generations model. Using data from Understanding Society and the National Child Development Study, we show that genetic endowments related to educational attainment and risk tolerance predict entrepreneurial selection and entrepreneurial earnings. We develop and estimate a model in which genetic endowments affect pre-market skill formation, preferences, and occupation-specific returns, while parents influence children through genetic transmission, investments, and transfers. We find that genetic heterogeneity accounts for a substantial share of intergenerational persistence (roughly 20–35 percent across earnings, consumption, and income ranks) and plays an important role in the formation of entrepreneurial dynasties. At the individual level, genetic endowments explain a modest but economically meaningful share of lifetime earnings variation, with effects that are twice as large among entrepreneurs as among wage workers. Finally, we compare financing wealth through credit expansion with financing talent through venture-capital intermediation. Financing talent improves mobility and aggregate output but increases inequality, and it is more efficient at boosting output than financing wealth.

JEL: D31, L26, J24, J62, E24

Keywords: entrepreneurship, genetics, intergenerational mobility, wealth distribution

*This paper builds on a chapter of Qianyu Yang's PhD dissertation. We thank Francesco Agostinelli, Naoki Aizawa, Benjamin Arnold, Minji Bang, Chao Fu, Diego Daruich, Luca Flabbi, Matthew Elliott, Hanming Fang, Eric French, Kai Liu, Dirk Krueger, Arnaud Maurel, Nicholas Papageorge, Sang Yoon (Tim) Lee, Ananth Seshadri, Christopher Taber, Constantine Yannelis, Ping Wang and Lichen Zhang as well as seminar and conference participants at the Econometric Society Asian Winter Meeting 2025, ESSGN 2025, Econometric Society North American Summer Meeting 2026, BSE Summer Forum 2026, Econometric Society Asia Meeting–China 2026, University of Cambridge, University of Edinburgh, University of Vienna, University of Pennsylvania, University of Wisconsin–Madison, Duke University, Johns Hopkins University, CUHK, HKUST, Tsinghua University, and Renmin University of China for helpful comments. This research received no external financial support. The authors have no relevant financial relationships or other potential conflicts of interest to disclose.

†Faculty of Economics, University of Cambridge. Email: qy254@cam.ac.uk

‡Faculty of Business and Economics, The University of Hong Kong. Email: njguo@hku.hk

§Faculty of Economics, University of Cambridge. Email: wz301@cam.ac.uk

1 Introduction

Entrepreneurship is central to understanding intergenerational mobility. It can provide a route out of inherited economic position when individuals with high-return ideas overcome financing constraints, but it can also reproduce advantage when business ownership, skills, networks, and capital access persist within families. A large literature documents strong intergenerational transmission of entrepreneurship. Existing explanations emphasize the transmission of financial resources, human capital, and family networks.[1] Yet entrepreneurial persistence is not fully accounted for by education, wealth, and observable skills (Levine and Rubinstein, 2017b). This residual persistence suggests that additional forms of latent heterogeneity may shape who selects into entrepreneurship, who remains successful, and how entrepreneurial advantage is transmitted across generations.

One important and long-standing mechanism underlying this intergenerational persistence is the genetic link between parents and children. While the role of innate ability has long been recognized in economics, it has historically been difficult to measure directly and therefore remains an "elephant in the room" in explaining entrepreneurial dynasties and intergenerational mobility.[2] Recent advances in molecular genetics now make this mechanism empirically accessible by providing a direct measure of genetic endowment, offering new insights into intergenerational transmission. Polygenic scores, indices that aggregate the small effects of thousands of genetic variants, serve as economically meaningful proxies for traits linked to education, risk tolerance, and other characteristics relevant to entrepreneurial behavior (Rietveld et al., 2013; Okbay et al., 2016; Lee et al., 2018; Hill et al., 2019; Okbay et al., 2022). These measures allow us to directly observe the genetic endowments of both parents and children, opening the door to a systematic analysis of how biological transmission contributes to entrepreneurial selection and intergenerational mobility.

This paper studies how genetic endowments shape entrepreneurship and intergenerational mobility. We begin with empirical evidence linking polygenic scores for educational attainment and risk tolerance to entrepreneurial intentions, entry, and occupation-specific earnings. We then develop a structural overlapping-generations model in which genetic endowments affect pre-market type formation, preferences, and occupation-specific earnings, while entrepreneurship is an endogenous occupational choice subject to financing frictions. The model allows us to quantify how genetic heterogeneity contributes to entrepreneurial dynasties, intergenerational mobility, and wealth concentration, and to study how alternative financing institutions allocate capital between wealth and entrepreneurial talent.

We first conduct a systematic empirical analysis of a broad set of polygenic scores to identify which dimensions of genetic endowment are most relevant for entrepreneurial behavior.

[1] See, for example, (Holtz-Eakin et al., 1994b,a; Hout and Rosen, 1999; Dunn and Holtz-Eakin, 2000; Fairlie and Robb, 2007; Lindquist et al., 2015).

[2] Early evidence from twin and family studies suggests that entrepreneurial propensity has a substantial heritable component; see, for example, Nicolaou and Shane (2010) and Lindquist et al. (2015). However, these approaches identify heritability without directly observing the underlying traits.

Polygenic scores for educational attainment and risk tolerance emerge as the most informative predictors, measured using PGS_{edu} from Okbay et al. (2022) and PGS_{risk} from Karlsson Linnér et al. (2019).[3] To move beyond correlations, our within-family specification exploits Mendelian randomization (Kong et al., 2018; Brumpton et al., 2020; Young et al., 2022; Biroli et al., 2026): conditional on the parents' polygenic scores, a child's deviation from the parental average reflects the random segregation of alleles at conception and is therefore orthogonal to family background and rearing environment, isolating the causal effect of genetic disposition on entrepreneurial choice. The reduced-form evidence points to distinct roles for the two scores. PGS_{risk} is more strongly associated with entrepreneurial intentions and entrepreneurial earnings, while PGS_{edu} is more closely related to realized entry and earnings in paid employment. These occupation-specific gradients motivate the two-dimensional genetic endowment used in the structural model.

In the model, individuals make consumption, saving, and occupational choices, while parents also choose investments in and transfers to their children. Genetic endowments affect outcomes through three channels. First, they shape the formation of pre-market types, which is also affected by parental genes and investments in the childhood period. Second, pre-market types affect preferences and productivity for each occupation, as well as risk preference, influencing selection into entrepreneurship and entrepreneurial performance. Third, conditional on pre-market type, genetic endowments directly shift earnings in each occupation. The model incorporates three channels through which parents affect the outcomes of their children, including genetic transmission, parental investment, and financial transfers.

We estimate the model using the simulated method of moments, combining two complementary longitudinal datasets: Understanding Society and the National Child Development Study. The former provides rich information on labor market outcomes and entrepreneurial activity, including a subsample of parent–child pairs with genetic data available for both generations. The latter offers detailed information on parental investments, childhood skill formation, and genetic data for children.

The model matches a rich set of targeted moments, including entrepreneurship shares, worker and entrepreneur earnings profiles across genetic endowments, parental investment, transfers, savings, and the Gini coefficients for income and wealth. It also replicates several untargeted moments, including intergenerational elasticities of earnings and consumption and a realistic wealth distribution across top wealth shares. These fits provide out-of-sample support for the model's core mechanisms.

Our estimation shows that genetic endowments for educational attainment and risk tolerance shape entrepreneurship through multiple channels. First, they affect the formation of pre-market types, which carry different preferences: high types are less risk-averse and derive higher non-pecuniary value from entrepreneurship. Second, genetic endowments generate occupation-

[3]Consistent with the literature, we define entrepreneurs as individuals running a business or professional practice, or partners in one—a group we label incorporated entrepreneurs—distinguishing them from the broader self-employed; legal incorporation status is not directly observed.

specific earnings premia. On the one hand, risk-related endowments are especially rewarded in entrepreneurship; on the other hand, education-related genetic endowments affect employee earnings, which in turn help entrepreneurs overcome borrowing constraints. Together, these channels shape both selection into entrepreneurship and performance conditional on entry.

Based on estimation results, we use the model to quantify the role of genetic heterogeneity through a sequence of counterfactual economies. First, we shut down the intergenerational genetic link, so that children draw genetic endowments and pre-market types independently of their parents'. This removes dynastic genetic transmission while preserving individual-level genetic heterogeneity. Second, we remove genetic heterogeneity altogether by shutting down genetic inputs in both type formation and earnings. This decomposition separates two genetic forces. The first is the effect of the intergenerational genetic link: because parents with favorable genetic endowments tend to have children with similar genes and also provide more investments and transfers, the previous generation's genetic advantages are passed on and shape the current generation's outcomes. The second is the effect of individual genetic heterogeneity within a generation: even when endowments are drawn independently of parents, individual genetic differences alone generate dispersion in outcomes.

The first main finding is that genetic heterogeneity is quantitatively important for understanding entrepreneurial dynasties and intergenerational mobility. Removing genetic mechanisms reduces persistence in earnings, consumption, and income ranks while increasing upward mobility. Across these measures, genetic heterogeneity accounts for roughly 20–35 percent of benchmark intergenerational persistence. Both dynastic transmission and individual genetic differences contribute to this effect. Together, these results provide a structural account of the genetic component of entrepreneurial transmission identified by adoption and family studies that decompose persistence into nature and nurture (Lindquist et al., 2015; Hvide and Oyer, 2018).

The second main finding concerns the role of genetic heterogeneity in shaping individual lifetime outcomes. Genetic endowments explain about 8.6 percent of the variation in lifetime earnings in the population as a whole, compared with roughly 14 percent among entrepreneurs and 7 percent among wage workers. They also explain about 5.5 percent of the variation in lifetime utility. These results indicate that genetic heterogeneity contributes meaningfully to differences in long-run economic outcomes, particularly among entrepreneurs. More broadly, the findings suggest that even the two polygenic scores we measure account for a modest but economically meaningful share of the persistent heterogeneity that structural life-cycle models use to explain long-run differences in outcomes (Huggett et al., 2011; Keane and Wolpin, 1997; Todd and Zhang, 2020). Because these two scores capture only a subset of the genetic variation relevant to economic behavior, the full contribution of genetic endowments is likely larger.

The third main finding is that genetic heterogeneity contributes to wealth concentration primarily through entrepreneurship, with effects that are concentrated in the right tail of the wealth distribution while leaving the main body largely unchanged. Genetic differences affect entrepreneurial productivity and earnings, and these differences compound through capital accumulation and intergenerational transmission. As a result, genetic heterogeneity helps explain

the emergence and persistence of large fortunes, with its strongest effects at the very top of the distribution. More broadly, our results suggest that measured genetic endowments account for part of the persistent entrepreneurial ability heterogeneity that structural models rely on to generate a fat right tail of the wealth distribution (Quadrini, 1999; Cagetti and De Nardi, 2006; De Nardi, 2004).

The fourth main finding concerns the design of entrepreneurial finance. We compare two approaches to expanding entrepreneurial capital: financing wealth through broad credit relaxation and financing talent through venture-capital intermediation. The two policies provide the same aggregate amount of entrepreneurial financing but allocate capital differently. Credit relaxation expands borrowing through collateral and therefore favors entrepreneurs with greater wealth. Venture-capital finance allocates capital based on expected entrepreneurial returns and therefore directs resources toward more productive entrepreneurs. We find that financing talent generates larger gains in upward mobility and aggregate output. However, it also increases income and wealth inequality because the additional returns accrue disproportionately to the most productive entrepreneurs. These results highlight a fundamental tradeoff between mobility, efficiency, and inequality in the design of entrepreneurial finance. More broadly, they suggest that mobility and inequality need not move in opposite directions when inequality reflects differences in entrepreneurial returns rather than unequal access to capital (Aghion et al., 2019; Jones and Kim, 2018; Hassler et al., 2007).

Related literature This paper contributes to three strands of literature.

First, our paper connects to the structural entrepreneurship literature. A robust empirical finding is the strong intergenerational persistence of entrepreneurship (Lindquist et al., 2015; Hvide and Oyer, 2018). Classic evidence highlights human capital and financial capital as important drivers, since parental self-employment and wealth strongly predict entrepreneurial entry among children (Holtz-Eakin et al., 1994b,a; Dunn and Holtz-Eakin, 2000; Fairlie and Robb, 2007; Guo and Leung, 2021). Yet these explanations do not fully account for sources of persistence. Even after conditioning on wealth, networks, and observed skills, substantial variation remains (Levine and Rubinstein, 2017b). Adoption studies provide especially compelling evidence: both pre-birth genetic inheritance and post-birth rearing environments shape entrepreneurial transmission (Lindquist et al., 2015). However, this literature typically treats heterogeneity as exogenous and does not model its intergenerational origins (Hamilton, 2000; Hamilton et al., 2019; Guo and Leung, 2021). Our contribution is to provide a structural framework in which inherited genetic endowments shape both entrepreneurial selection and performance, and to quantify how biological transmission contributes to entrepreneurial dynasties.

Second, the paper contributes to the emerging literature integrating molecular genetic measures into economics. Polygenic scores have been shown to predict education, income, and wealth (Rietveld et al., 2013; Okbay et al., 2016; Lee et al., 2018; Okbay et al., 2022). Recent work highlights the importance of gene–environment interaction and genetic nurture in shaping economic outcomes (Kong et al., 2018; Barth et al., 2020; Houmark et al., 2024; Arold et al.,

2025; Biroli et al., 2026). However, without an explicit causal framework, correlations between polygenic scores and economic outcomes are difficult to interpret (Durlauf and Rustichini, 2023). Structural modeling is therefore essential to quantify how genes operate through economic channels.

The closest papers to ours are Barth et al. (2022a) and Rustichini et al. (2023). Rustichini et al. (2023) shows that polygenic scores predict educational attainment and intergenerational mobility. Barth et al. (2022a) incorporates genetic endowments into a life-cycle model of consumption, savings, and portfolio choice. We build on this work by placing genetic endowments inside an explicit intergenerational model of occupational choice. This allows genetic endowments to affect mobility through type formation, preferences, occupation-specific earnings, parental investment, and financial transfers. The model therefore connects measured genetic heterogeneity to the economic mechanisms through which entrepreneurship either disrupts or reinforces intergenerational persistence.

Third, our paper contributes to the literature on intergenerational mobility. A large empirical literature documents strong persistence in income, education, occupation, and wealth across generations (Solon, 1992; Mazumder, 2005; Chetty et al., 2014; Gregg et al., 2017). A parallel structural literature studies how parental investments, borrowing constraints, human capital formation, and bequests shape mobility (Becker and Tomes, 1979, 1986; Solon, 2004; Lee and Seshadri, 2019; Abbott et al., 2019; Daruich, 2018). In these models, persistent unobserved heterogeneity is often important for matching the data, but its source is usually left implicit. Our paper provides a measurable source of such heterogeneity and studies how it interacts with entrepreneurship to shape intergenerational mobility.

This mobility perspective also changes how we interpret wealth concentration. Entrepreneurship can be a mobility channel when high-return individuals from less wealthy families receive capital, but it can also reinforce persistence when entrepreneurial ability, resources, and transfers align within dynasties. This connects the mobility literature to work showing that entrepreneurship is central to wealth concentration and the upper tail of the distribution (Quadrini, 1999; Cagetti and De Nardi, 2006; De Nardi, 2004). Our model captures both forces by allowing genetic heterogeneity to affect entrepreneurial selection, persistence, and returns. The results show that these mechanisms shape intergenerational mobility, while their implications for wealth are uneven: modest for the main body of the distribution and concentrated at the upper tail.

The remainder of the paper is organized as follows. Section 2 presents empirical evidence on the relationship between genetic endowments and entrepreneurial outcomes. Section 3 develops the overlapping-generations model. Section 4 describes the estimation strategy, identification, and model fit. Section 5 presents the key structural estimates and the channels linking genetic heterogeneity to entrepreneurship. Section 6 quantifies the importance of genetic heterogeneity for intergenerational mobility, individual lifetime outcomes, and wealth concentration. Section 7 compares financing wealth with financing talent. Section 8 concludes.

2 Genetic Endowments and Entrepreneurship: Empirical Evidence

Our empirical analysis combines two complementary longitudinal datasets: Understanding Society (the UK Household Longitudinal Study) and the National Child Development Study (NCDS). Together, these datasets allow us to link genetic endowments to entrepreneurial choices, family environments, human capital accumulation, and lifetime economic outcomes.

Understanding Society is our primary source for studying entrepreneurship. The survey is a nationally representative household panel that builds on the British Household Panel Survey (BHPS, 1991–2008) and has been fielded annually since 2009, covering approximately 40,000 households drawn from a stratified, clustered probability sample of the UK residential population. It provides rich annual information on employment, income, education, health, attitudes, and detailed occupational histories. A key advantage of the survey is its ability to distinguish individuals who run a business or professional practice, or who are partners in a business, from other forms of self-employment. We classify the former group as entrepreneurs and, following the terminology commonly used in the entrepreneurship literature, refer to them as incorporated entrepreneurs. This classification is intended to capture business-owning and firm-building activity, although legal incorporation status is not directly observed. This distinction is important because such business owners are more likely to engage in innovation and firm growth, whereas unincorporated self-employment often reflects necessity or flexible labour-supply decisions (Hurst and Pugsley, 2011; Levine and Rubinstein, 2017a).

We classify individuals running a business or professional practice, or who are partners in such a practice, as entrepreneurs, and we group full-time employees together with the unincorporated self-employed (sole proprietors, freelancers, and subcontractors) as workers. This classification follows the entrepreneurship literature, which views unincorporated self-employment as economically closer to wage employment than to firm-building activity. Observations with incomplete employment information or part-time status are excluded (see Appendix B.2).

A second advantage of Understanding Society is the availability of genetic data. Of the 9,919 genotyped individuals, 7,029 have complete information on education, employment, and wealth-related outcomes and form our main analysis sample. In addition, the survey includes 1,299 individuals for whom we can impute the genetic endowments of both parents, allowing us to study intergenerational channels such as genetic nurture and assortative mating. Appendix Table A.1 summarizes the genetic measures used in the analysis.

The NCDS complements Understanding Society by providing a detailed view of early-life environments and skill formation. The study follows all children born in Britain during a single week of March 1958 and contains rich information on parental investments, school quality, family background, cognitive skills, and non-cognitive traits. Later waves track educational attainment, occupations, and earnings throughout adulthood. Genetic data collected through the age-44/45 Biomedical Survey allow us to construct polygenic scores and examine how genetic endowments interact with family inputs over the life cycle.

The two datasets play distinct but complementary roles. Understanding Society provides detailed evidence on entrepreneurial choices and career dynamics, while the NCDS sheds light on the developmental mechanisms linking genetic endowments, parental investments, and human capital accumulation. The NCDS does not, however, collect the occupational and tax-based information needed to identify incorporated entrepreneurship, so it cannot on its own support the entrepreneurial classification central to our analysis; its comparative advantage lies instead in its uniquely rich, prospectively measured record of early-life skills, parental investments, and family environments. We therefore rely on Understanding Society to define and study entrepreneurial outcomes, and on the NCDS to trace the developmental pathways through which genetic differences translate into those outcomes.

We proceed in three steps. Section 2.1 identifies the genetic dimensions most strongly associated with entrepreneurship. Section 2.2 documents how these genetic endowments relate to entrepreneurial entry, human capital accumulation, and family background. Finally, Section 2.3 exploits within-family variation to isolate the causal effects of genetic endowments from environmental confounders.

2.1 Which Genetic Endowments Matter for Entrepreneurship

We measure genetic endowments at the molecular level using polygenic scores (PGSs), which aggregate information across many genetic variants to predict a range of outcomes, following the recent genoeconomics literature.[4] Because entrepreneurship is not a standard phenotype in existing large-scale GWAS datasets and reflects both innate predispositions and post-birth environmental influences, there is currently no validated polygenic score designed specifically for entrepreneurial choice. We therefore begin by identifying the genetic traits that distinguish entrepreneurs from wage workers and examining which dimensions of genetic variation are most strongly associated with entrepreneurial entry.

Combining annual career histories with genetic data, we find that entrepreneurs exhibit systematically different genetic profiles from workers.[5] Appendix Figure A.1 shows the PGSs that differ significantly between the two groups. Measured in standard deviation units relative to the population mean, entrepreneurs have, on average, a 17% higher PGS_{edu} and a 13% higher PGS_{risk}. Several other PGSs, including those related to health, personality, and income, also differ across groups, although the magnitudes are substantially smaller.

To quantify these relationships, we estimate the association between each PGS and the probability of ever becoming an entrepreneur using Understanding Society, with the following logistic specification:

$$\Pr(\text{Entrepreneur}_i = 1) = \Lambda\left(\alpha_0 + \alpha_1 \text{PGS}_i + \alpha_2' \text{PC}_i + \alpha_3 \text{SEX}_i + \epsilon_i\right) \tag{2.1}$$

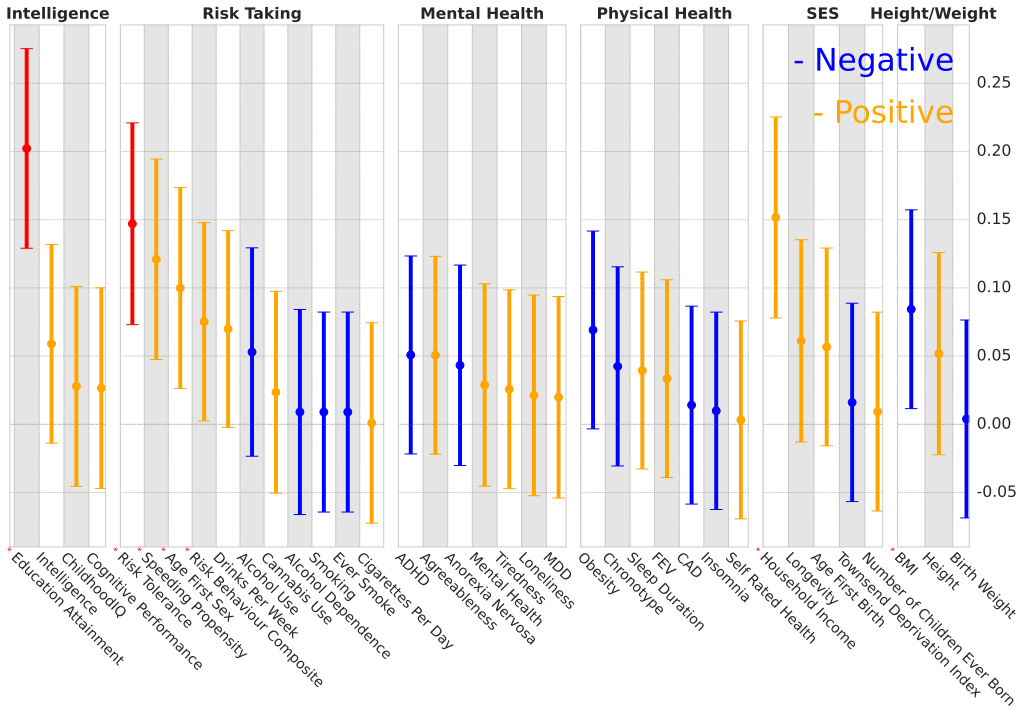
[4] Detailed information on the construction and validation of these genetic measures is provided in Appendix A.

[5] The entrepreneur group includes individuals ever observed running their own business; the remainder are classified as workers.

where Entrepreneur$_i$ is an indicator for whether individual i has ever been an entrepreneur. The coefficient α_1 captures the association between a one-standard-deviation increase in the polygenic score and the log-odds of entrepreneurial entry, conditional on gender (SEX$_i$) and the top ten genetic principal components (PC$_i$). These principal components account for population stratification (systematic genetic differences across ancestry groups that may correlate with socioeconomic factors and otherwise bias estimates) (Price et al., 2006).

Figure 1 summarizes the estimated associations.[6] Two polygenic scores stand out. PGS_{edu} (Okbay et al. 2022) is strongly positively associated with entrepreneurship, consistent with the importance of human capital and cognitive skills in venture creation. PGS_{risk} (Karlsson Linnér et al. 2019) is also a strong predictor, reflecting the central role of willingness to bear uncertainty in entrepreneurial decisions.

Figure 1: Associations Between PGSs and the Probability of Ever Being an Entrepreneur



Notes: Standardized coefficients are estimated from separate logistic regressions of ever being an entrepreneur on individual polygenic scores (PGSs). Each regression controls for gender and the top 10 genetic principal components. PGSs are grouped into six domains: intelligence, mental health, risk-related behavior, physical health, socioeconomic status (SES), and height/weight. Each bar displays the absolute value of the estimated association, with horizontal lines indicating 95% confidence intervals. Yellow bars denote positive associations and blue bars negative ones. The red bar highlights PGS_{edu}. Statistically significant estimates are marked with red asterisks. Abbreviations: MDD = major depressive disorder; CAD = coronary artery disease; FEV = forced expiratory volume (lung function).

For the rest of the paper, we focus on these two dimensions of genetic endowment. Empirically, they exhibit the largest and most robust associations with entrepreneurial entry. Theoretically, they correspond closely to two key determinants of entrepreneurship emphasized in the literature:

[6]We enter each PGS separately because the scores are genetically correlated and each is a noisy proxy for its underlying trait; a single-score specification recovers the marginal association of each genetic predisposition with entrepreneurship, which is the object of interest here. We report a joint specification in Appendix C.1 as a robustness check.

human capital and risk preferences (Kihlstrom and Laffont, 1979; Koudstaal et al., 2016; Levine and Rubinstein, 2018). While other polygenic scores display statistically significant correlations, their economic interpretation is less direct and their explanatory power is substantially weaker. We therefore define genetic endowment using these two dimensions. The next section documents the descriptive patterns associated with these endowments (Section 2.2), before isolating their causal genetic component using a within-family design (Section 2.3).

2.2 Descriptive Associations Between Genetic Endowments and Entrepreneurship

We next examine how the two genetic dimensions identified above relate to entrepreneurship, human capital accumulation, and family background. To do so, we divide individuals into four groups according to whether their PGS_{edu} and PGS_{risk} lie above or below the sample median. Table 1 reports descriptive statistics for these groups, combining entrepreneurial outcomes from Understanding Society with measures of childhood skills, parental investments, and family background from the NCDS. Panel A focuses on entrepreneurship and earnings, Panel B on pre-market skills and parental investments, Panel C on family background, and Panel D on parental genetic endowments. These patterns are descriptive and motivate the causal analysis developed in Section 2.3.

Entrepreneurship and Earnings. Panel A reveals a clear gradient in entrepreneurial activity. The probability of ever becoming an entrepreneur rises monotonically with both genetic dimensions, from 10.3% among individuals with low values of both scores to 13.3% among those with high values of both. The same pattern appears for earnings: individuals with higher PGS_{edu} and PGS_{risk} earn substantially more on average. These patterns suggest that both dimensions are associated not only with entrepreneurial selection but also with broader economic success.

Human Capital and Parental Investment. Panel B shows that differences in genetic endowments are already reflected in pre-market skills and preferences.[7] Individuals with high PGS_{edu} obtain university degrees at much higher rates and exhibit substantially stronger cognitive skills by age 16. Differences in risk-taking behavior align closely with variation in PGS_{risk}, consistent with the interpretation of this score as capturing predispositions toward risk-taking.[8] Parental investments also vary systematically across genetic groups. Using standardized measures of parental time investment at ages 7, 11, and 16, we find that children with higher PGS_{edu} receive greater parental investments throughout childhood.[9] This pattern is consistent with parental

[7] Appendix B.1 describes the NCDS measures of cognitive skills, non-cognitive skills, parental time investment, and school quality.

[8] Risk taking is measured using the question: "Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?" The answer ranges from 1 (unwilling to take risks) to 10 (fully prepared to take risks).

[9] Parental time investment is not observed directly. At each age the NCDS records a set of indicators, such as the frequency of outings with the child, whether the parents read to the child, and the teacher's assessment of

Table 1: Descriptive Statistics by PGS Group

	PGS_{edu}			
	Low (< 50%)		High (> 50%)	
PGS_{risk}	Low	High	Low	High
Panel A: Individual characteristics				
Ever became entrepreneur (prop.)	0.103	0.111	0.122	0.133
Earnings (2015 £)				
Mean	18,418	18,971	21,245	22,204
Std. dev.	14,722	16,151	17,646	18,841
Panel B: Parental investments and pre-market outcomes				
Pre-market outcomes				
University degree (prop.)	0.270	0.312	0.430	0.471
Risk-taking (0–10 scale)	4.896	5.242	5.041	5.315
Cognitive skills at 16 (std.)	−0.257	−0.283	0.265	0.275
Non-cognitive skills at 16 (std.)	−0.031	−0.205	0.154	0.085
Parental investments (std.)				
Investment at age 7	−0.169	−0.147	0.177	0.136
Investment at age 11	−0.161	−0.182	0.169	0.171
Investment at age 16	−0.146	−0.156	0.145	0.156
Panel C: Family background, measured at age 14				
Father's characteristics				
University degree (prop.)	0.047	0.058	0.096	0.115
Entrepreneur (prop.)	0.040	0.045	0.043	0.059
Mother's characteristics				
University degree (prop.)	0.031	0.042	0.049	0.078
Entrepreneur (prop.)	0.004	0.007	0.006	0.008
Parents' household income (2015 £)*	18,862	18,580	19,063	19,516
Panel D: Parental genes				
PGS^{p}_{edu} (std.)	−0.282	−0.271	0.296	0.277
PGS^{p}_{risk} (std.)	−0.109	0.143	−0.108	0.135

Notes: This table presents descriptive statistics for individuals grouped by median splits of their polygenic scores (PGSs) for educational attainment (PGS_{edu}) and risk preference (PGS_{risk}). Earnings are annual gross income in 2015 British pounds. Variables labeled "(prop.)" are proportions. Variables labeled "(std.)" are standardized to have a mean of 0 and a standard deviation of 1 in the full sample. Parental genes (PGS^{p}) are the standardized average of both parents' scores. Parental investments represent a standardized index of time investment derived from teacher and parent reports. * indicates metrics from the NCDS sample; others are from Understanding Society.

inputs reinforcing, rather than offsetting, differences in children's endowments. As a result, genetic differences are associated with widening gaps in human capital accumulation over the life cycle.

Family Background and Gene–Environment Correlations. Panel C documents a strong relationship between genetic endowments and family background. Individuals with high PGS_{edu}

parental interest in the child's education. Following the human-capital measurement literature (Cunha et al., 2010; Attanasio et al., 2020), we treat these as noisy measures of a latent parental-investment factor, which we estimate separately at each age and standardize to mean zero and unit variance.

are substantially more likely to have highly educated parents and to grow up in families with entrepreneurial experience. For example, the share of fathers holding a university degree is more than twice as large among individuals in the high-PGS_{edu} groups as among those in the low-PGS_{edu} groups. Similar gradients are evident for parental entrepreneurship.

These patterns highlight an important challenge for interpreting cross-sectional associations between genes and entrepreneurship. Genetic endowments are not randomly assigned across family environments. Instead, they are correlated with parental human capital, occupational choices, and other family characteristics that also influence entrepreneurial outcomes.

Intergenerational Transmission of Genetic Endowments. Panel D shows a strong intergenerational correlation in polygenic scores. Individuals with high values of PGS_{edu} and PGS_{risk} tend to have parents with similarly high scores, reflecting the direct transmission of genetic endowments across generations. This pattern highlights why genetic differences and family-background differences are closely intertwined in the data.

Taken together, the evidence suggests that PGS_{edu} and PGS_{risk} may influence entrepreneurial outcomes through multiple channels. These genetic endowments are associated not only with entrepreneurial entry itself, but also with human capital accumulation, parental investments, and broader family environments. As a result, simple cross-sectional correlations cannot distinguish the role of genetic endowments from that of the environments in which they are expressed.

Biological and adopted children. To shed light on the importance of genetic transmission, we exploit the distinction between biological and adopted children. If genetic factors contribute to occupational persistence, parent-child similarity should be stronger among biological children, who share both genes and family environment, than among adopted children, who share only the family environment. Consistent with this prediction, having a parent in the large-employer class increases a biological child's probability of reaching the same class by 2.4 percentage points, whereas the corresponding effect for adopted children is statistically indistinguishable from zero.[10] The difference between the two estimates is statistically significant. Thus, the persistence observed among biological children largely disappears among adopted children raised in similar households, pointing to an important genetic component in the intergenerational transmission of occupational status. Appendix C.2 provides details of the empirical design, estimating equation, and results across occupational classes.

2.3 Within-Family Evidence on Entrepreneurial Intention, Entry, and Earnings

The descriptive patterns in Section 2.2 reveal strong associations between genetic endowments, entrepreneurship, and family background. However, individuals with different polygenic scores

[10]The large-employer class is an official Understanding Society category, comprising individuals who are self-employed with more than 25 employees.

also grow up in systematically different environments. Children with high PGS_{edu}, for example, tend to have more educated parents, receive greater parental investments, and benefit from more favorable socioeconomic circumstances. As a result, cross-sectional correlations cannot distinguish the effect of an individual's own genetic endowment from the influence of family background.

To address this issue, we exploit within-family genetic variation. The key idea is that children inherit a random combination of alleles from their parents. Conditional on parental genotype, the deviation of a child's polygenic score from the parental average arises from Mendelian segregation and is therefore plausibly unrelated to family-level environmental factors. This variation allows us to isolate the component of genetic endowment that is orthogonal to parental background and genetic nurture.

Our analysis focuses on entrepreneurial intention, entrepreneurial entry, and occupation-specific earnings. We use the 1,299 individuals for whom mid-parental polygenic scores can be constructed from observed biological relatives using the parental-genotype imputation procedure of Young et al. (2022).[11] Appendix B.3 provides details of the construction procedure.

For each trait ($k \in \{\text{edu}, \text{risk}\}$), let $(PGS_{k,i})$ denote the individual's polygenic score and let $(\overline{PGS}^{p}_{k,i})$ denote the corresponding mid-parental score. We decompose the individual's score into a parental component and a Mendelian component:

$$D_{k,i} \equiv PGS_{k,i} - \overline{PGS}^{p}_{k,i},$$

where $D_{k,i}$ captures the child's deviation from the parental average. This deviation reflects the random component of genetic inheritance and serves as our primary source of identification.

We estimate

$$\begin{aligned} y_i = \; & \alpha + \theta_{\text{edu}} D_{\text{edu},i} + \theta^{p}_{\text{edu}} \overline{PGS}^{p}_{\text{edu},i} \\ & + \theta_{\text{risk}} D_{\text{risk},i} + \theta^{p}_{\text{risk}} \overline{PGS}^{p}_{\text{risk},i} + \gamma' Z_i + \varepsilon_i. \end{aligned} \tag{2.2}$$

where (Z_i) includes demographic controls, the top ten genetic principal components, and, in richer specifications, family socioeconomic controls (parental occupational class at child's age 14). The coefficients (θ_{edu}) and (θ_{risk}) capture the association between outcomes and the child's Mendelian component, while $(\theta^{p}_{\text{edu}})$ and $(\theta^{p}_{\text{risk}})$ capture the association with parental genetic endowments. Because the Mendelian component reflects the random segregation of alleles during meiosis, it is independent of parental genotype. These estimates are substantially less exposed to confounding from shared family background and genetic nurture than conventional cross-sectional associations (Kong et al., 2018; Young et al., 2022). Throughout, we interpret the estimates as reduced-form evidence on the channels through which genetic endowments shape entrepreneurial outcomes.

This parameterization has two additional advantages. First, it nests the standard individual-PGS specification. Imposing the restriction $\theta_k = \theta^{p}_{k}$ collapses equation (2.2) to a regression of

[11] The 1,299 individuals are on average younger than the full sample of 9,919 individuals. Appendix C.4 uses entropy balancing (Hainmueller, 2012) to reweight this subsample to match the full-sample distribution of age, sex, years of education, and risk preference, and shows that the main results are robust to this adjustment.

outcomes on the individual's PGS together with the same control vector Z_i. The individual-PGS specification can therefore be viewed as a restricted version of equation (2.2). Second, the Mendelian component is theoretically orthogonal to the parental mean, providing a convenient framework for addressing measurement error in genetic variables. Appendix A.1 exploits this property to implement an error-in-variables correction. In the main text we focus on OLS estimates and use the EIV estimates as a robustness check.

Table 2 reports estimates for entrepreneurial intention and entrepreneurial entry. Columns (1)–(3) show that entrepreneurial intention is primarily associated with risk-related genetic endowments.[12] A one-standard-deviation increase in the child's PGS_{risk} deviation raises the probability of expecting to start a business by approximately 1.2–1.7 percentage points, and the estimate remains stable after controlling for family socioeconomic status. By contrast, the coefficient on the child's PGS_{edu} deviation is small and statistically insignificant throughout. The mid-parental PGS coefficients are also insignificant, suggesting that entrepreneurial aspirations are linked to idiosyncratic variation in risk-related genetic endowments rather than family-level influences captured by parental genetic endowments.

Table 2: Reduced-Form Evidence: Entrepreneurial Intention and Entrepreneurial Entry

	Entrepreneurial Intention			Entrepreneurial Entry		
	(1)	(2)	(3)	(4)	(5)	(6)
Child's PGS_{edu} deviation	0.006	0.011	0.010	0.022***	0.019*	0.017*
	(0.006)	(0.007)	(0.007)	(0.008)	(0.010)	(0.010)
Child's PGS_{risk} deviation	0.012*	0.017**	0.017**	0.004	0.003	0.003
	(0.007)	(0.008)	(0.008)	(0.008)	(0.010)	(0.010)
Mid-parental PGS_{edu}	–	−0.000	−0.002	–	0.026**	0.023**
		(0.009)	(0.008)		(0.011)	(0.011)
Mid-parental PGS_{risk}	–	0.006	0.004	–	0.006	0.004
		(0.010)	(0.010)		(0.011)	(0.011)
Family SES controls	–	–	✓	–	–	✓
N	1,276	1,276	1,276	1,299	1,299	1,299

Notes: This table reports OLS reduced-form estimates using a linear probability model. In columns (1)–(3), the dependent variable is a binary indicator for whether the individual ever responds "Yes" to "Do you expect to start your own business?" in any survey wave. In columns (4)–(6), the dependent variable is a binary indicator for whether the individual ever becomes an entrepreneur. Child's PGS deviation refers to $D_{k,i} = PGS_{k,i} - \overline{PGS}^{p}_{k,i}$. Default controls include sex, the top ten genetic principal components, survey-wave fixed effects, and missing dummies. Family SES controls comprise parental occupational-class indicators. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Columns (4)–(6) turn to realized entrepreneurial entry. Here the dominant pattern shifts from PGS_{risk} to PGS_{edu}. A one-standard-deviation increase in the child's PGS_{edu} deviation

[12]The question "Do you expect to start your own business?" is fielded in alternating BHPS waves in 1992–2002 and in all available Understanding Society waves from wave 1 (2009/10) onward. In the full genotyped sample, 4.9% of respondents ever answer yes; in the reduced-form analysis subsample, 5.7%. Individuals with no valid response in any wave are excluded from the entrepreneurial-intention specifications, reducing the sample from 1,299 to 1,276.

increases the probability of entrepreneurial entry by approximately 1.6–2.2 percentage points across specifications, with little sensitivity to the inclusion of family socioeconomic controls. The mid-parental PGS_{edu} is also positive and statistically significant, implying that both individual-specific and family-level education-related endowments are associated with entrepreneurial entry. In contrast, neither the child's nor the parental PGS_{risk} exhibits a statistically significant relationship with entry. Taken together, the results suggest that risk-related endowments primarily influence entrepreneurial aspirations, whereas education-related endowments are more important in translating those aspirations into realized entrepreneurial activity.

We next examine earnings across occupations. Table 3 shifts attention from occupational entry to occupational returns. Columns (1)–(3) focus on entrepreneurs. A one-standard-deviation increase in the child's PGS_{risk} deviation is associated with approximately 17–24 percent higher entrepreneurial earnings, and the estimate remains statistically significant after controlling for family socioeconomic background. Thus, among individuals from comparable family environments, those with higher risk-related genetic endowments earn substantially more in entrepreneurship. In contrast, the child's PGS_{edu} deviation has no robust association with entrepreneurial earnings.

Table 3: Reduced-Form Evidence: Log Earnings

	Entrepreneur Log Earnings			Worker Log Earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
Child's PGS_{edu} deviation	−0.010	−0.016	0.001	0.083***	0.071***	0.068***
	(0.064)	(0.084)	(0.083)	(0.006)	(0.008)	(0.007)
Child's PGS_{risk} deviation	0.167**	0.236***	0.234***	0.005	0.012	0.011
	(0.073)	(0.075)	(0.075)	(0.007)	(0.008)	(0.008)
Mid-parental PGS_{edu}	–	0.020	0.073	–	0.101***	0.092***
		(0.104)	(0.122)		(0.009)	(0.009)
Mid-parental PGS_{risk}	–	−0.052	−0.037	–	−0.007	−0.012
		(0.118)	(0.122)		(0.010)	(0.010)
Family SES controls	–	–	✓	–	–	✓
N	432	432	432	8,290	8,290	8,290

Notes: This table reports OLS reduced-form estimates for log annual gross earnings. The sample is restricted to 1,299 individuals with mid-parental PGS constructed from observed biological relatives; the earnings regressions further restrict to ages 24–60. In columns (1)–(3), the dependent variable is entrepreneur log annual earnings. In columns (4)–(6), the dependent variable is worker log annual earnings. Child's PGS deviation refers to $D_{k,i} = PGS_{k,i} - \overline{PGS}^{p}_{k,i}$. Default controls include sex, the top ten genetic principal components, age, and age squared. Family SES controls comprise parental occupational-class indicators. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Columns (4)–(6) present the corresponding estimates for workers. In contrast to the entrepreneurial sample, the dominant pattern is the effect of the child's PGS_{edu} deviation. A one-standard-deviation increase in PGS_{edu} is associated with roughly 6–8 percent higher wage earnings, consistent with the well-established role of education-related genetic endowments in educational attainment and labor-market productivity (Bolt et al., 2025). The mid-parental

PGS_{edu} also has a positive and significant association with earnings, indicating that family-level factors linked to educational endowments continue to matter for wage determination. By contrast, the child's PGS_{risk} deviation has little association with wage earnings, and the sizable entrepreneurial risk premium observed in columns (1)–(3) is largely absent among workers.

The earnings results therefore reveal clear occupation-specific gradients. Risk-related endowments are most strongly associated with success in entrepreneurship, whereas education-related endowments are more strongly associated with earnings in wage employment. These patterns remain robust after correcting for sample selection using entropy balancing (Hainmueller, 2012) and for measurement error using the error-in-variables (EIV) procedure of Heckman et al. (2013). Detailed robustness results are reported in Appendix C.3 and Appendix C.4.

Taken together, the within-family estimates reveal distinct roles for the two genetic dimensions. Risk-tolerance endowments are most strongly associated with entrepreneurial intentions and entrepreneurial earnings, while educational-attainment endowments are associated with entrepreneurial entry and wage earnings. Combined with the descriptive evidence in Section 2.2, these findings motivate the key mechanisms of the model. Genetic endowments influence occupation-specific returns, interact with parental investments in shaping pre-market skills, and are transmitted across generations through genetic inheritance. Section 3 formalizes these channels within an overlapping-generations framework of occupational choice, discretizing each of the two genetic dimensions into high and low categories.

3 Model

The model consists of two main building blocks. First, the skill formation process captures the interplay between nature and nurture during childhood, following Houmark et al. (2024). Skill heterogeneity is summarized by pre-market types, which determine initial labor-market productivity upon entry at age 24. Second, we incorporate entrepreneurship decisions within a general equilibrium overlapping-generations (OLG) framework, following Cagetti and De Nardi (2006, 2009), allowing individuals to self-select into different career paths based on inherited wealth and multidimensional abilities. The formal specification appears in the subsequent subsections.

3.1 Model Setup

Environment and Timeline The economy is modeled in a dynastic overlapping-generations framework where each period j spans six years. The life cycle consists of three active stages from age 24 to 60, followed by retirement (Figure 2). Before age 24, individuals are dependent children living in their parents' households, and all decisions are made by parents. State variables associated with children are denoted by the superscript c, while those without superscripts refer to parents.

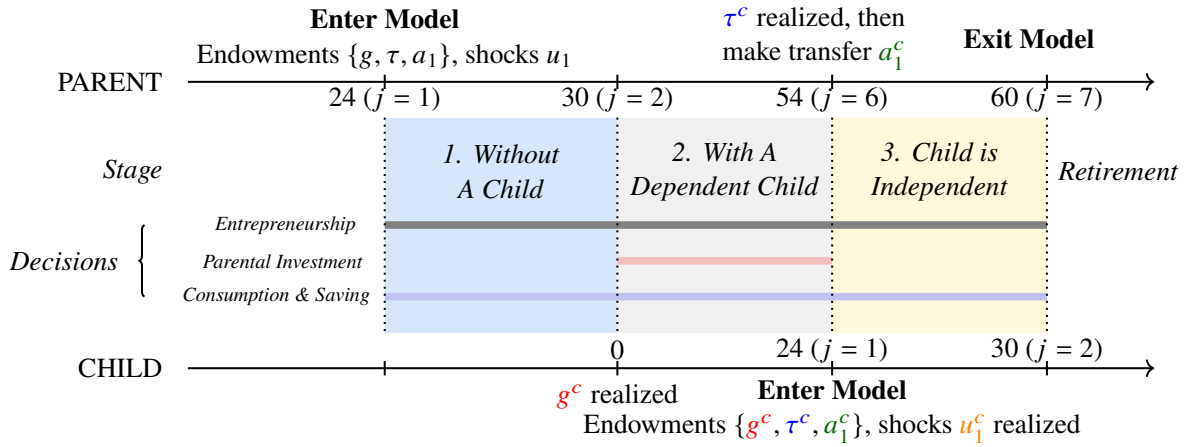
Agents enter the model at age 24 ($j = 1$) as independent adults, characterized by genetic

endowment g, accumulated human capital τ, and inherited assets a_1. In this stage, they observe their productivity shock u_1 and decide on consumption, savings, and occupation (entrepreneurial decision).

At age 30 ($j = 2$), agents enter parenthood. The child's genetic endowment g^c is realized, and the household expands to include a dependent child.[13] Between ages 30 and 48 ($j = 2$–5), in addition to consumption, saving, and occupational choices, parents allocate time to parental investment that shapes the child's skill formation.

At age 54 ($j = 6$), the child reaches adulthood. Their pre-market type τ^c and initial productivity shock u_1^c are realized. Observing these, the parent then makes a one-time *financial* transfer a_1^c to launch the child's independent household, remains active for one final period, and then retires at age 60.

Figure 2: Model Timeline



Preferences Households derive utility from consumption, the non-pecuniary benefits of entrepreneurship, and parental time investment. The per-period utility of an individual of age j and type τ is given by

$$U(c_j, \omega_j; \tau) = \frac{c_j^{1-\gamma_\tau}}{1-\gamma_\tau} + b_\tau \, I[\omega_j = E] - \xi \, t_j^p, \tag{3.1}$$

where c_j denotes consumption, ω_j denotes occupational choice, and t_j^p denotes parental time investment in children.

The first term is a standard CRRA utility function with type-specific risk aversion parameter γ_τ. Allowing γ_τ to vary across pre-market types captures systematic differences in willingness to bear consumption risk associated with educational attainment and risk preferences. The second term, b_τ, represents the non-pecuniary utility from entrepreneurship. This component is also

[13] We assume a single child born at a deterministic period to maintain computational tractability, a common simplification in the quantitative OLG literature (e.g., Daruich, 2018; Lee and Seshadri, 2019). While we recognize that endogenizing the fertility decision or allowing for multiple children born at different stages of the life cycle would provide richer dynamics, we leave these as further extensions.

type-specific and captures preferences for autonomy and independence that may lead individuals to pursue entrepreneurship even when its expected financial return is lower than that of paid employment (Hamilton, 2000; Hurst and Pugsley, 2011; Koudstaal et al., 2016).

The final term captures the utility cost of parental time investment. The variable t_j^p measures the time devoted by parents to child development, including educational activities, supervision, and other forms of parental engagement. Unlike monetary expenditures, this input represents the opportunity cost of parental time. In the model, parental time investment is undertaken only during the parenting periods of the life cycle, when it contributes to children's skill formation; specifically, $t_j^p > 0$ for $j \in \{2, 3, 4, 5\}$ and $t_j^p = 0$ in all other periods. The parameter ξ governs the utility cost of providing this time input. Parental time investment t^p is distinct from the monetary *financial* transfer a_1^c that parents make when the child reaches independence at $j = 6$ (defined in equation (3.10)): t^p is a time input provided during the parenting periods that shapes the child's pre-market type through the formation function (3.2), whereas a_1^c is a financial transfer that augments the child's initial assets. The two thus operate as separate channels of intergenerational influence.

Genetics We focus on two dimensions of genetic endowment, measured by polygenic scores: educational attainment (PGS_{edu}) and risk tolerance (PGS_{risk}). Following Barth et al. (2022b) and to maintain computational tractability, we discretize each score at the sample median into High (H) and Low (L) categories. This discretization also allows us to remain agnostic about whether individuals observe their exact polygenic scores: agents are assumed not to know their precise genetic endowments, but to infer their coarse type (high versus low) and condition decisions on this information. This yields a state space of four composite genetic types:

$$g = (PGS_{edu} \otimes PGS_{risk}) \in \{(\mathrm{H}, \mathrm{H}),\ (\mathrm{H}, \mathrm{L}),\ (\mathrm{L}, \mathrm{H}),\ (\mathrm{L}, \mathrm{L})\}.$$

The intergenerational transmission of these traits is governed by a Markov transition matrix, $G(g^c \mid g)$, estimated directly from parent-child pairs. This matrix, visualized in Figure 3, encapsulates the dual nature of genetic inheritance. First, the diagonal elements capture *intergenerational persistence*: parents with favorable genetic endowments are statistically more likely to pass these traits to their offspring, creating a biological component of intergenerational persistence. Second, the off-diagonal elements capture the *randomness of the genetic transmission*: even parents with extreme genetic advantages face non-negligible probabilities of having children with different genetic profiles. This stochasticity is crucial, as it allows the model to distinguish between deterministic dynasty effects and probabilistic genetic transmission.

Pre-market Types and Their Formation We use pre-market types to summarize the heterogeneity individuals bring to the labor market at age 24. These types serve as a parsimonious representation of pre-market productivity and preferences, linking genetic and family background characteristics to subsequent educational, occupational, and entrepreneurial choices. While it would be desirable to model educational attainment and risk preferences as separate latent

dimensions, doing so would substantially increase the state space. Instead, we combine these dimensions into a single pre-market type, which preserves the key economic mechanisms while maintaining computational tractability. Following Bonhomme et al. (2019); Jolivet and Postel-Vinay (2025), we recover latent types using k-means clustering applied to observed educational attainment and risk tolerance.

Let $\tau^c \in \{L, H\}$ denote the child's realized pre-market type. Type formation depends on parental type τ, parental genetic endowment g, the child's genetic endowment g^c, and parental time investment t^p (Keane and Wolpin, 1997; Todd and Zhang, 2020). We model the probability that a child becomes the high type using a binary logit specification:

$$\Pr(\tau^c = H \mid \tau, g, g^c, t^p) = \Lambda(\eta), \qquad \Pr(\tau^c = L \mid \tau, g, g^c, t^p) = 1 - \Lambda(\eta), \tag{3.2}$$

where $\Lambda(x) = \exp(x)/(1 + \exp(x))$ is the logistic function and

$$\eta = \delta_0 + \delta_\tau + \delta_{g^c} + \delta_g + \delta^{t^p}_{g^c} \cdot t^p. \tag{3.3}$$

Here, δ_0 is a constant, while δ_τ, δ_{g^c}, and δ_g capture the effects of parental type, child genetic endowment, and parental genetic endowment, respectively, with the omitted category normalized to zero. The $\delta^{t^p}_{g^c}$ allows the productivity of parental investment to vary with the child's genetic endowment.

This specification captures several mechanisms emphasized in the nature–nurture literature (Cunha and Heckman, 2008; Houmark et al., 2024; Bolt et al., 2025; Biroli et al., 2026). Child genetic endowments g^c generate *direct genetic effects*, while parental endowments g capture *genetic nurture*, allowing parental genes to influence child outcomes through the family environment beyond what is biologically transmitted. Parental time investment t^p represents the environmental input into skill formation. Because parents choose t^p in response to parental and child characteristics, the model also accommodates gene-environment interactions arising from endogenous parental behavior.

The interaction term $\delta^{t^p}_{g^c} \cdot t^p$ introduces *gene-environment interaction* ($G \times E$): it allows the return to parental investment to differ across genetic endowments. If the return to investment is higher for high-endowment children, this implies a reinforcement effect. By modeling both endogenous parental investment and heterogeneous returns to that investment, the framework allows genetic and environmental factors to interact through parental behavior as well as through the skill-formation technology itself (Biroli et al., 2026).

Production Function Our specifications allow for two primary channels through which genetic endowments influence individual productivity. First, genes affect the pre-market types, which in turn influence occupation-specific productivity. Second, conditional on the pre-market type, we allow genes to have a direct impact on individual productivity. This direct effect captures the role of innate ability differences emphasized in the existing literature (Lee and Seshadri, 2019).

We specify the efficiency units of ability for occupation $\omega \in \{E, W\}$ (entrepreneur E; worker

W) as a function of pre-market heterogeneity (τ), genetic endowments (g), and an idiosyncratic productivity shock ($u_{\omega,j}$). Formally, the log efficiency units at age j are given by:

$$\ln f_{\omega,j} = \beta_{0,\omega} + \beta_{g,\omega} + \beta_{\tau,\omega} + u_{\omega,j}, \qquad \omega \in \{E, W\}. \tag{3.4}$$

The terms $\beta_{g,\omega}$ and $\beta_{\tau,\omega}$ are occupation-specific fixed effects indexed by the discrete genetic type g and pre-market type τ, respectively.

$$u_{\omega,j} \;=\; \rho_\omega\, u_{\omega,j-1} \;+\; \eta_{\omega,j}, \qquad \eta_{\omega,j} \sim \mathcal{N}(0, \sigma^2_\omega). \tag{3.5}$$

The term $u_{\omega,j}$ captures persistent unobserved human capital and career-specific advantages. For regular workers, $u_{W,j}$ represents "dining table" human capital (informal information and professional norms transmitted within the family environment). For entrepreneurs, $u_{E,j}$ captures specific business ideas and access to family-based social networks within the business world. To capture the intergenerational transmission of these unobserved advantages, we assume that when children reach independence, they inherit their initial productivity level $u^c_{\omega,1}$ from their parent's current career shock $u_{\omega,6}$ at the same calendar period following an AR(1) process:

$$u^c_{\omega,1} = \rho^c_\omega u_{\omega,6} + \epsilon_\omega, \omega \in \{E, W\} \tag{3.6}$$

where ρ^c_ω governs the strength of the intergenerational correlation in unobserved ability. This assumption is economically intuitive as it recognizes that a parent's professional success and accumulated knowledge create a localized informational environment that benefits the child at the start of their own career.

Worker income is determined by the market wage rate w and the individual's efficiency units:

$$y_{W,j} \;=\; w\, f_{W,j}. \tag{3.7}$$

Entrepreneurial earnings are determined by a static capital choice problem. Entrepreneurs operate a decreasing-returns-to-scale technology and choose capital k to maximize profit, subject to a credit constraint arising from market incompleteness:[14]

$$y_{E,j} \;=\; \max_{0 \le k \le \bar{a}_j} \left\{ P\, f_{E,j}\, k^\nu - (r + \delta)\, k \right\}, \tag{3.8}$$

where P is aggregate entrepreneurial productivity, $\nu \in (0, 1)$ is the span-of-control parameter (capital elasticity), r is the risk-free rate, and δ is the depreciation rate. The maximum investable capital, $\bar{a}_j$, is determined by a leverage constraint following Evans and Jovanovic (1989) and

[14]The closed-form solution for the optimal unconstrained capital investment is $k^* = \left(\frac{r+\delta}{\nu P f_{E,j}}\right)^{\frac{1}{\nu-1}}$. With the constraint, the optimal capital is $k = \min\{k^*, \bar{a}_j\}$.

Kiyotaki and Moore (1997):

$$\bar{a}_j - a_j \leq \chi a_j \quad \Longrightarrow \quad \bar{a}_j \leq (1+\chi) a_j, \tag{3.9}$$

where a_j denotes the agent's net worth and $\chi \geq 0$ is the parameter governing the maximum leverage ratio.

Comparing equation (3.7)) with equation (3.8)), a fundamental difference emerges: while worker productivity is linear in human capital, entrepreneurial productivity can be scaled up through capital investment. Consequently, an individual's decision to enter entrepreneurship over regular wage work depends not only on innate talent, encapsulated by genetic endowments and pre-market types, but also critically on their wealth. Since credit constraints limit the scalability of a business, the choice is heavily influenced by dynastic family support via financial transfers a_1^c.

Budget Constraint Financial markets are incomplete. Households smooth consumption by trading a risk-free, non-state-contingent asset with return r. However, they face a strict liquidity constraint that precludes borrowing ($a_{j+1} \geq 0$).

The household's budget constraint evolves over the life cycle to account for specific milestones. We model parental time investment as a utility cost incurred during the parenthood stage ($2 \leq j \leq 5$): in each of these periods, the parent allocates a fraction t_j^p of their time to child-rearing, with the magnitude of this cost scaled by ξ. The committed investment level $t^p \in [0, 1]$ is chosen at $j = 2$ and sustained thereafter, so that $t_j^p = t^p$ for $2 \leq j \leq 5$ and $t_j^p = 0$ otherwise. The expenditure side varies by age: standard savings occur in early adulthood ($1 \leq j \leq 5$), the intergenerational transfer a_1^c is made at the beginning of $j = 6$, and the saving for retirement a_{retire} occurs at $j = 6$ after the transfer decision.

The resulting budget constraint is given by:

$$(1+r)a_j + y_{\omega,j} = \begin{cases} c_j + a_{j+1} & \text{if } 1 \leq j \leq 4, \\ c_j + a_{j+1} + a_1^c & \text{if } j = 5, \\ c_j + a_{\text{retire}} & \text{if } j = 6. \end{cases} \tag{3.10}$$

where c_j denotes household consumption. On the income side, $y_{\omega,j}$ represents the potential market earnings from occupation ω. On the expenditure side, a_1^c represents the one-time financial transfer made to the child at the beginning of $j = 6$, and a_{retire} denotes the terminal savings carried into the retirement phase.

3.2 Recursive Formulation

We formulate the household's life-cycle optimization problem recursively. Time is discrete and indexed by the agent's age j. Each period follows a specific sequence of events and decisions where, at the beginning of each period j, the household observes a vector of predetermined

state variables S_j. The composition of S_j evolves over the life cycle: (1) Independent young adulthood ($j = 1$): $S_1 = \{g, \tau, a_1\}$. (2) The parenthood phase ($2 \leq j \leq 5$): $S_j = \{g, \tau, a_j, g^c\}$, where g^c is the child's realized genetic endowment. (3) Household splitting and independent child ($j = 6$): $S_6 = \{g, \tau, a_6, g^c, \tau^c\}$, incorporating the child's realized pre-market type.

The decision-making sequence within each period j proceeds in three distinct steps: (1) Realization of shocks: The agent observes idiosyncratic productivity shocks $u_j = (u_{W,j}, u_{E,j})$ for both potential occupations. (2) Occupational selection: Based on the current state S_j and shocks u_j, the household chooses its occupation $\omega_j \in \{E, W\}$ (entrepreneurship or wage work) to maximize the expected present value of lifetime utility. (3) Continuous decisions: Conditional on the chosen occupation and the resulting income, the household makes continuous decisions regarding consumption c_j, savings a_{j+1}, and, during the parenthood stage, the level of parental investment t^p. Prior to the child's independence, the parent also chooses a cash transfer a_1^c, which becomes the child's initial asset.

To aid the notation, we introduce the ex-ante value function $V_j(S_j, u_j)$, which is defined as the maximum over the occupation-specific value functions:

$$V_j(S_j, u_j) = \max \left\{ V_j^E(S_j, u_j),\, V_j^W(S_j, u_j) \right\},$$

where $V_j^\omega(S_j, u_j)$ represents the value of choosing occupation ω and behaving optimally thereafter.

Pre-Parenthood ($j = 1$): In period $j = 1$ (age 24), the agent enters the model as an independent young adult. The household is currently childless, but the agent anticipates the formation of a family in the subsequent period ($j = 2$). The state vector is given by $S_1 = (g, \tau, a_1)$, where g denotes the agent's genetic endowment, τ is the fixed pre-market type, and a_1 represents initial assets.

The agent observes the current vector of productivity shocks u_1 and solves for the optimal occupation ω, consumption c_1, and savings a_2. In doing so, the agent forms expectations over two sources of future uncertainty: the evolution of their own income process (u_2) and the realization of the child's genetic endowment (g^c), which occurs at the onset of the next period.

The value function is:

$$V_1(S_1, u_1) = \max_{\omega \in \{E,W\}} \left\{ V_1^\omega = \max_{c_1} \left\{ U(c_1; \omega, \tau) + \beta \sum_{g^c} G(g^c \mid g) \int_{u_2} V_2(S_2, u_2)\, dF(u_2 \mid u_1) \right\} \right\}$$

where $G(g^c \mid g)$ represents the intergenerational genetic transition probability defined in the Genetics section. The summation over g^c captures the agent's anticipation of their future child's genetic traits, which will become a state variable in period $j = 2$.

The Parenthood Phase ($j = 2, \ldots, 5$): Upon entering parenthood ($j = 2$), the household structure expands to include a dependent child. The parent now cares not only for their own consumption and occupational success but also for the future human capital of their child. The

state vector expands to include the child's genetic endowment, which is realized at the onset of this stage. The state is given by $S_j = (g, \tau, a_j, g^c)$, where g^c denotes the child's genetic endowment.

Investment Commitment and Early Parenting ($j = 2$): In the first period of child-rearing, the parent observes the child's genetic endowment g^c and chooses a committed level of time investment $t^p \in [0, 1]$ that will be sustained throughout the child's dependency. This captures the following trade-off: increasing parental investment t^p enhances the probability of the child acquiring a high-productivity type τ^c (via the type-formation function in equation (3.2)), but reduces the parent's utility.

The value function for this stage is:

$$V_2(S_2, u_2) = \max_{\omega \in \{E,W\}} \left\{ V_2^\omega = \max_{\{c_2, t^p\}} \left\{ U(c_2; \omega, \tau) + \beta \int_{u_3} V_3(S_3, u_3; t^p)\, dF(u_3 \mid u_2) \right\} \right\}$$

Intermediate Parenting Periods ($j = 3 - 4$): In the intermediate parenting periods, the level of parental investment t^p is no longer a choice variable but a fixed obligation pre-determined at $j = 2$. It enters the utility function as a disutility of parenting, imposing a utility loss in each parenting period. The state vector remains $S_j = (g, \tau, a_j, g^c)$, and the value function is:

$$V_j(S_j, u_j; t^p) = \max_{\omega \in \{E,W\}} \left\{ V_j^\omega = \max_{c_j} \left\{ U(c_j; \omega, \tau) + \beta \int_{u_{j+1}} V_{j+1}(S_{j+1}, u_{j+1}; t^p)\, dF(u_{j+1} \mid u_j) \right\} \right\}$$

Final Parenting and Type Realization ($j = 5$): Period $j = 5$ is the final stage of parenting. At the end of this period, the child's pre-market type τ^c will be realized. The parent forms expectations over this realization based on the formation function $\Pr(\tau^c \mid S_5, t^p)$ following equation 3.2, which depends on the cumulative effect of parental investment, family background, and genetic endowments.

The value function incorporates this transition into the next generation's independence:

$$V_5(S_5, u_5; t^p) = \max_{\omega \in \{E,W\}} \left\{ V_5^\omega = \max_{c_5} \left\{ U(c_5; \omega, \tau) + \beta \int_{u_6} \sum_{\tau^c} \tilde{V}_6(\tilde{S}_6, u_6) \cdot \Pr(\tau^c \mid S_5, t^p)\, dF(u_6 \mid u_5) \right\} \right\}$$

Here, $\tilde{V}_6(\cdot)$ denotes the joint continuation value at the beginning of period $j = 6$, when the child becomes independent. The state vector $\tilde{S}_6$ collects the parent's and child's states prior to separation.

Transition to Independence and Intergenerational Transfers ($j = 6$): The child becomes independent at this stage (age 24). The household's optimization problem unfolds in two sequential steps. First, the parent determines the intergenerational wealth transfer when their child becomes independent. Second, conditional on this transfer, the parent makes final labor supply and retirement saving decisions.

Optimal financial Transfers: Upon the realization of the child's pre-market type τ^c and initial productivity shocks u_1^c, the parent solves a capital allocation problem to determine the optimal *financial* transfer, a_1^c. This decision is driven by the trade-off between the parent's remaining life-cycle needs (retirement) and their altruistic concern for the child's initial wealth. The chosen transfer a_1^c acts as the decision variable that decomposes $\tilde{S}_6$ into separate state vectors for the parent (S_6) and the independent child (S_1^c):

$$\underbrace{(g, \tau, \tilde{a}_6, g^c, \tau^c)}_{\tilde{S}_6} \xrightarrow{a_1^c} \begin{cases} S_6 = (g, \tau, \underbrace{\tilde{a}_6 - a_1^c}_{a_6}) & \text{(Parent's Post-Transfer State)} \\ S_1^c = (g^c, \tau^c, \underbrace{a_1^c}_{\text{Child's Initial Asset}}) & \text{(Child's Initial State)} \end{cases}$$

where $\tilde{a}_6$ denotes the total consolidated wealth available to the household at period $j = 6$.

Formally, the parent maximizes a weighted sum of their own continuation value and the child's expected value:

$$\tilde{V}_6(\tilde{S}_6, u_6) = \max_{0 \le a_1^c \le \tilde{a}_6} \left\{ V_6(S_6, u_6) + \lambda\beta \int_{u_1^c} V_1(S_1^c, u_1^c)\, dF(u_1^c \mid u_6) \right\}$$

The parameter λ governs the degree of altruism. The integral captures the expectation over the child's initial productivity shocks u_1^c following the AR(1) process specified in equation (3.5). We allow u_1^c to be correlated with the parent's current shock u_6, reflecting the intergenerational persistence of unobserved ability beyond what is explained by genetics and pre-market types.

Late-Career Decisions and Preparation for Retirement: Following the transfer, the household reverts to a single-agent structure. The parent's remaining assets are $a_6 = \tilde{a}_6 - a_1^c$. The agent then solves a standard consumption-savings and occupational choice problem for this final active period, knowing that the next stage is retirement.

The value function is given by:

$$V_6(S_6, u_6) = \max_{\omega \in \{E,W\}} \left\{ V_6^\omega = \max_{c_6} \{U(c_6; \omega, \tau) + \phi(a_{\text{retire}}; \tau)\} \right\}$$

Here, $\phi(a_{\text{retire}}; \tau)$ captures a terminal value representing the utility derived from consumption during the retirement phase (ages 60+), which depends on the retirement savings a_{retire} and the agent's type. Details of its closed-form derivation can be found in Appendix D.1.

3.3 Equilibrium

A stationary recursive competitive equilibrium consists of a set of prices $\{w, r\}$, a set of household decision rules, and a stationary probability measure $m^*(S, u)$ over the state space, such that:

Individual Optimality: Given factor prices (w, r), the decision rules for occupation ω, consumption c, savings a', and capital investment k solve the household's dynamic programming problems defined in the Recursive Formulation section.

Corporate Production: There exists a representative non-entrepreneurial (corporate) firm that produces output Y using a standard Cobb–Douglas technology:

$$Y = AK_C^{\alpha} L_C^{1-\alpha},$$

where K_C is corporate capital and L_C is corporate labor. Profit maximization implies that factor prices equal their marginal products:

$$w = (1-\alpha)A\left(\frac{K_C}{L_C}\right)^{\alpha}, \qquad r = \alpha A\left(\frac{K_C}{L_C}\right)^{\alpha-1} - \delta.$$

Labor Market Clearing: The demand for labor by the corporate sector equals the aggregate supply of efficiency units from households choosing wage work:

$$L_C = \int \mathbb{I}_{\{\omega(S,u)=W\}} \, f_W(S,u) \, dm^*(S,u),$$

where $\mathbb{I}_{\{\cdot\}}$ is the indicator function for the occupational choice.

Capital Market Clearing: The total supply of financial assets from all households equals the sum of capital demanded by the corporate sector and the entrepreneurial sector:

$$K_C = \int a(S,u) \, dm^*(S,u) \; - \; \int \mathbb{I}_{\{\omega(S,u)=E\}} \, k(S,u) \, dm^*(S,u).$$

Here, the first integral represents the aggregate net worth of the economy, and the second integral captures the capital employed privately by entrepreneurs.

Stationary Distribution: The distribution m^* is invariant, satisfying:

$$m^*(\mathcal{S}, \mathcal{U}) = \int P((S,u), (\mathcal{S}, \mathcal{U})) \, dm^*(S,u)$$

for all Borel sets $(\mathcal{S}, \mathcal{U})$, consistent with the optimal decision rules and the exogenous stochastic processes for shocks and genetic transmission.

4 Estimation, Identification and Model Fit

This section describes how the model is disciplined by the data. Our estimation strategy combines calibration, reduced-form estimation, and structural estimation. We proceed in three steps. First,

we calibrate a small set of standard parameters, including the interest rate, depreciation rate, and credit constraint, using values commonly employed in the literature. Second, we estimate key components of the state space outside the structural model, namely the latent pre-market types and the intergenerational genetic transmission process. Using Understanding Society, we recover latent type classifications based on educational attainment and risk preferences and estimate the transition matrices governing the transmission of genetic endowments across generations. Third, we estimate the remaining structural parameters using the Simulated Method of Moments (SMM). For each candidate parameter vector, we solve the model for its stationary equilibrium, simulate the implied distributions of earnings, wealth, occupations, and intergenerational outcomes, and choose the parameter values that minimize the distance between model-implied and empirical moments.

A key advantage of this approach is that it allows the model to be disciplined using moments drawn from multiple datasets. Childhood investment moments are obtained from the NCDS, while life-cycle earnings, wealth, occupational choices, and intergenerational outcomes are drawn from Understanding Society. As summarized in Table 4, combining these complementary sources of information enables us to identify both the early-life mechanisms governing type formation and the life-cycle processes that shape earnings, entrepreneurship, and wealth accumulation.

When forming the weighting matrix, we do not use the fully optimal weighting matrix. Instead, we weight each moment by the inverse of its sampling variance. Del Boca et al. (2014) shows that this estimator remains consistent in large samples. Relative to computing the optimal weighting matrix, this approach substantially reduces computational burden. Moreover, Altonji and Segal (1996) argues that efficiency gains from the optimal weighting matrix may be limited in finite samples.

4.1 Pre-set Parameters

We set the annual risk-free interest rate at 4.69%, following the long-run evidence in Jordà et al. (2019). The capital depreciation rate is calibrated at 6% annually, consistent with Cagetti and De Nardi (2006). Following Evans and Jovanovic (1989), we impose a credit constraint that limits total entrepreneurial capital investment to 1.22 times the entrepreneur's net worth. These parameter values align with established estimates in the structural entrepreneurship literature and provide a conservative baseline for our analysis.

4.2 Estimation Outside the Model

To preserve tractability while retaining the key economic channels, we summarize this heterogeneity using a discrete set of latent types identified via k-means clustering, following Bonhomme et al. (2019) and Jolivet and Postel-Vinay (2025). This approach offers two advantages. First, by clustering individuals to minimize within-group dispersion in education and risk preferences, the procedure maps observed cross-sectional variation into a small number of empirically grounded types consistent with the model structure. Second, because the clustering is implemented outside

Table 4: Model Parameters, Identification Strategy, and Data Sources

Parameter	Description	Targeted Empirical Moments	Source
Panel A: Calibrated Parameters following the literature			
r	Annual risk-free interest rate	4.69%	Jordà et al. (2019)
δ	Depreciation rate	6%	Cagetti and De Nardi (2006)
χ	Credit constraint	0.22	Evans and Jovanovic (1989)
Panel B: Parameters Estimated Outside the Model (First-stage)			
$G(g^c \mid g)$	Genetic transmission matrix	Markov matrix of PGS from parent to child	Understanding Society
Panel C: Parameters Estimated via SMM (Second-stage)			
Pre-market Type Formation			
$\Pr(\tau^c \mid \cdot)$	Type formation function	Probability of child being high pre-market type, conditional on parental type, parental gene group (8 moments) & Parental time investment received by child's gene group (4 moments)	Understanding Society & NCDS
Worker Sector Technology			
$f_{j,W}(\cdot)$	Worker ability profile	Mean worker earnings by gene group (4 moments) & pre-market type (2 moments)	Understanding Society
w	Wage per efficiency unit	Mean worker earnings (1 moment)	Understanding Society
ρ_w, σ_w	Worker income shock	Autocorrelation & Std. dev. of worker earnings (2 moments)	Understanding Society
Entrepreneurial Sector Technology			
$f_{j,E}(\cdot)$	Entrepreneurial ability profile	Mean entrep. earnings by gene group (4 moments) & pre-market type (2 moments)	Understanding Society
P	Entrepreneurial TFP	Mean entrepreneur earnings (1 moment)	Understanding Society
ν	Return to capital	Income Gini (1 moment)	WID
ρ_E, σ_E	Entrepreneur income shock	Autocorrelation & Std. dev. of entrepreneur earnings (2 moments)	Understanding Society
Intergenerational Transmission of Income Shocks			
$F(u_1^c \mid u_6)$	Shock transmission	Parents Occupation → Children Occupation (4 moments)	Understanding Society
Preferences and Utility			
β	Discount factor	Mean saving ratio (1 moment)	Understanding Society
γ_τ	Risk aversion	Saving ratio before retirement & Consumption growth volatility ratio across types (2 moments)	Understanding Society
λ	Altruism factor	Mean financial transfers (1 moment)	English Longitudinal Study of Ageing (ELSA)
b_τ	Non-pecuniary utility	Fraction of entrepreneurs by type and genes (6 moments)	Understanding Society
ξ	Utility cost of time investment	Mean parental investment (1 moment)	Understanding Society
Macro Aggregates			
A	Corporate TFP	Inverts firm FOCs given observed factor prices w, r and model-implied K_C, L_C	Understanding Society
α	Capital share		

Notes: This table summarizes the model parameters and their corresponding moments. Panel A lists parameters calibrated to standard values in the literature. Panel B lists parameters estimated directly from the data outside the model in a first step. Panel C lists parameters estimated via Simulated Method of Moments (SMM) by minimizing the distance between model-generated moments and data moments. The column "Targeted Empirical Moments" specifies which data moments primarily identify each parameter. WID: Global database of income and wealth inequality. NCDS: 1958 British birth cohort study. Understanding Society: Annual UK household panel survey.

the structural model, it substantially reduces computational burden while ensuring that the classification reflects meaningful patterns in the data.

Formally, we define a vector of individual-specific characteristics m_i for each individual $i \in \{1, \ldots, N\}$, comprising years of education and risk preference. The k-means algorithm partitions individuals into K latent types by solving:

$$\min_{\{p_i\} \in \{1,\ldots,K\}^N} \sum_{i=1}^{N} \|m_i - \bar{m}(p_i)\|^2,$$

where $\bar{m}(p_i)$ denotes the cluster-specific mean vector and $\|\cdot\|$ represents the Euclidean norm.

We recover two distinct types, summarized in Table 5.[15] "Low pre-market types" are characterized by lower educational attainment, lower risk taking and lower earnings, while "High pre-market types" display systematically higher outcomes across all three dimensions.

Table 5: Empirical Characteristics of Pre-market Types

Type	Education (years)	Risk Taking (0–10)	Earnings (2015 £)
Low pre-market type	17.13	4.4	14,184
High pre-market type	20.30	6.2	27,929

Notes: Means of education, risk-taking score, and earnings for the two latent types identified via k-means clustering on Understanding Society data.

Genetic Inheritance Process We estimate the genetic transmission process by constructing separate probability transition matrices for the intergenerational transmission of PGS_{edu} and PGS_{risk}. Estimating the two processes separately is supported by two pieces of empirical evidence. First, our analysis shows that the distribution of PGS_{risk} is statistically indistinguishable across the bottom and top halves of the PGS_{edu} distribution, and vice versa.[16] Second, genetic research indicates that educational attainment and risk tolerance are shaped by largely distinct polygenic architectures (Plomin and von Stumm, 2018; Visscher et al., 2017). Together, this evidence justifies modeling the transmission of PGS_{edu} and PGS_{risk} as independent processes, which substantially reduces computational complexity.

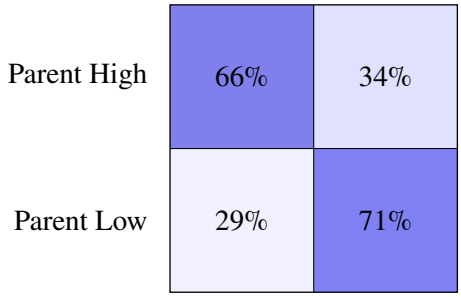
We estimate the transition matrices using parent–child genetic pairs from Understanding Society (see Appendix A.2 for details). Figure 3 presents the estimated transition probabilities.

Two key patterns emerge. First, both PGSs display strong intergenerational persistence: parents with high PGSs are significantly more likely to have children with high PGSs. Second, substantial randomness remains: parents with extreme PGS values still have non-negligible probabilities of having children with moderate or opposite PGS values. This reflects the genetic inheritance process, which combines persistence with randomness.

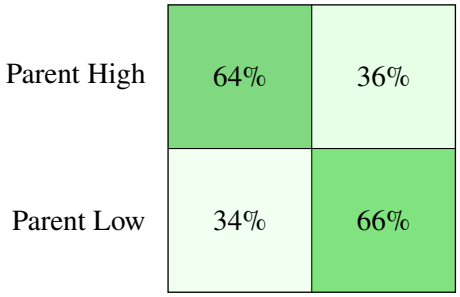
[15] Detailed identification results are provided in Appendix D.2.

[16] Figure A.2 in Appendix A.2 reports Kolmogorov–Smirnov tests comparing PGS distributions across splits of the other PGS. In both Understanding Society and the NCDS, we fail to reject the null of identical distributions.

Figure 3: Markov Transition Matrices of PGS (Parent → Child).



PGS_{edu} Transition Matrix (Parent → Child)



PGS_{risk} Transition Matrix (Parent → Child)

4.3 Estimation of Structural Parameters via SMM

We estimate the remaining structural parameters using the Simulated Method of Moments (SMM). Panel C of Table 4 summarizes the mapping between parameters and empirical moments. Identification follows a simple economic logic: each parameter is disciplined by the data feature it directly governs in equilibrium.

Pre-market Type Formation: We estimate the type formation process in Equation (3.2). The terms involving τ, g, and g^c are type-specific effects for discrete parental type, parental genetic type, and child genetic type, respectively. The equation links a child's realized type (τ^c) to their genetic endowments (g^c), parental type (τ), and parental investment (t^p); we construct the investment index from NCDS measures of time and school inputs. Identification rests on two sets of moments: the probability of high pre-market type classification by parental type and gene group (8 moments from Understanding Society), and the differential parental time investment received across child gene groups (4 moments from the NCDS).[17] Because parental and child genetic endowments are linked through the first-stage transition matrices, the child-gene effect δ_{g^c} is identified jointly with these moments rather than from independent variation in g^c.

Production and Earnings: The worker and entrepreneurial efficiency-unit functions ($f_{j,W}$, $f_{j,E}$) determine how the effective units of human capital supplied in each occupation vary across genetic groups and pre-market types. Mean earnings by gene group and pre-market type identify the level and slope of these profiles because changes in ability directly scale earnings conditional on occupation.

The income shock parameters (ρ_ω, σ_ω) are identified from earnings autocorrelations and cross-sectional dispersion. The persistence parameter ρ_ω governs how strongly current earnings predict future earnings in occupation ω, making it tightly linked to observed life-cycle autocorrelation.

[17] Given the limited number of observations for which parental genetic endowments, child genetic endowments, and child type are jointly observed, we do not directly target moments conditioning on all three variables. Instead, we combine moments on type formation by parental background with moments on parental investment by child genetic group, which provide more precisely estimated empirical discipline for the type-formation process.

The innovation standard deviation σ_ω controls the magnitude of earnings shocks, which maps directly into the cross-sectional variance of earnings.

The span-of-control parameter (ν) determines how entrepreneurial returns scale with capital and ability. A higher ν amplifies heterogeneity in entrepreneurial outcomes, increasing upper-tail dispersion. Because entrepreneurs concentrate at the top of the distribution, this dispersion passes through to the overall income Gini, which is therefore particularly informative about ν.

We identify ρ^c_ω using intergenerational occupational transitions. Intuitively, a higher ρ^c_ω implies that children of highly productive parents begin their careers with systematically higher latent productivity, increasing the probability of entering similar occupations and generating stronger parent–child earnings persistence. In the model, this parameter primarily governs the cross-generational correlation in occupational status and income ranks. We therefore discipline ρ^c_ω using moments that capture parent-to-child transitions in occupation, as reported in Understanding Society.

Preferences and Utility: The discount factor (β) governs intertemporal substitution and the accumulation of assets over the life cycle. A higher β increases the value of future consumption, leading households to save more and accumulate greater wealth. We therefore use the aggregate saving rate to discipline the level of patience in the economy.

Risk aversion (γ_τ) governs both precautionary saving and the extent to which households smooth consumption in response to income fluctuations. Higher risk aversion strengthens buffer-stock saving motives, increasing wealth holdings and reducing the transmission of income shocks to consumption. We use two sets of moments to identify γ_τ. First, the pre-retirement saving rate helps separate risk aversion from patience: conditional on β, a steeper accumulation profile prior to retirement implies greater precautionary saving and therefore a higher value of γ_τ (Gourinchas and Parker, 2002). Second, we exploit differences in consumption insurance across pre-market types using the ratio

$$\frac{\text{Var}(\log c \mid \tau = High)/\text{Var}(\log y \mid \tau = High)}{\text{Var}(\log c \mid \tau = Low)/\text{Var}(\log y \mid \tau = Low)}.$$

This moment identifies the differential $\gamma_H - \gamma_L$. A higher consumption-to-income volatility ratio indicates that consumption responds more strongly to income shocks, reflecting weaker precautionary smoothing and therefore lower risk aversion. Because high types are more risk tolerant, their consumption passes through income shocks more strongly, so the ratio exceeds one (Blundell et al., 2008).[18] The terminal value function $\phi(a_{\text{retire}}; \tau)$ depends on type and cash on hand and is derived in closed form in Appendix D.1, capturing precautionary saving motives in a tractable way.

The altruism parameter (λ) directly affects financial transfers. Because transfers are a forward-looking choice driven by dynastic preferences. Given measurement limitations in

[18]The data contain information on food expenditure rather than total consumption. We therefore use $\text{Var}(\log c_{\text{food}})$ as a proxy for $\text{Var}(\log c)$ and assume that the proportional relationship between food and total consumption volatility is similar across types, so that the cross-type ratio remains informative about relative consumption insurance.

financial-transfer data in Understanding Society, we instead discipline this moment using the mean financial transfer from the English Longitudinal Study of Ageing (ELSA), as reported by Bolt et al. (2024). These financial transfers include both inter vivos transfers received earlier in life and bequests received later in life. The amounts are discounted to age 24, the age at which transfers occur in the model.

The non-pecuniary utility of entrepreneurship (b_τ) captures occupational tastes beyond financial returns. Conditional on genes and types, expected profits alone cannot explain observed entry rates. The residual variation in entrepreneurial participation therefore identifies b_τ by reconciling observed occupational choices with predicted income differences.

The utility cost parameter (ξ) directly governs the disutility parents incur when allocating time to child investment. Higher ξ shifts the interior optimum toward lower time allocations, reducing the probability that the child attains a High pre-market type. Conditional on the estimated returns from the type formation stage, the mean parental investment observed in Understanding Society provides one moment that directly pins down ξ.

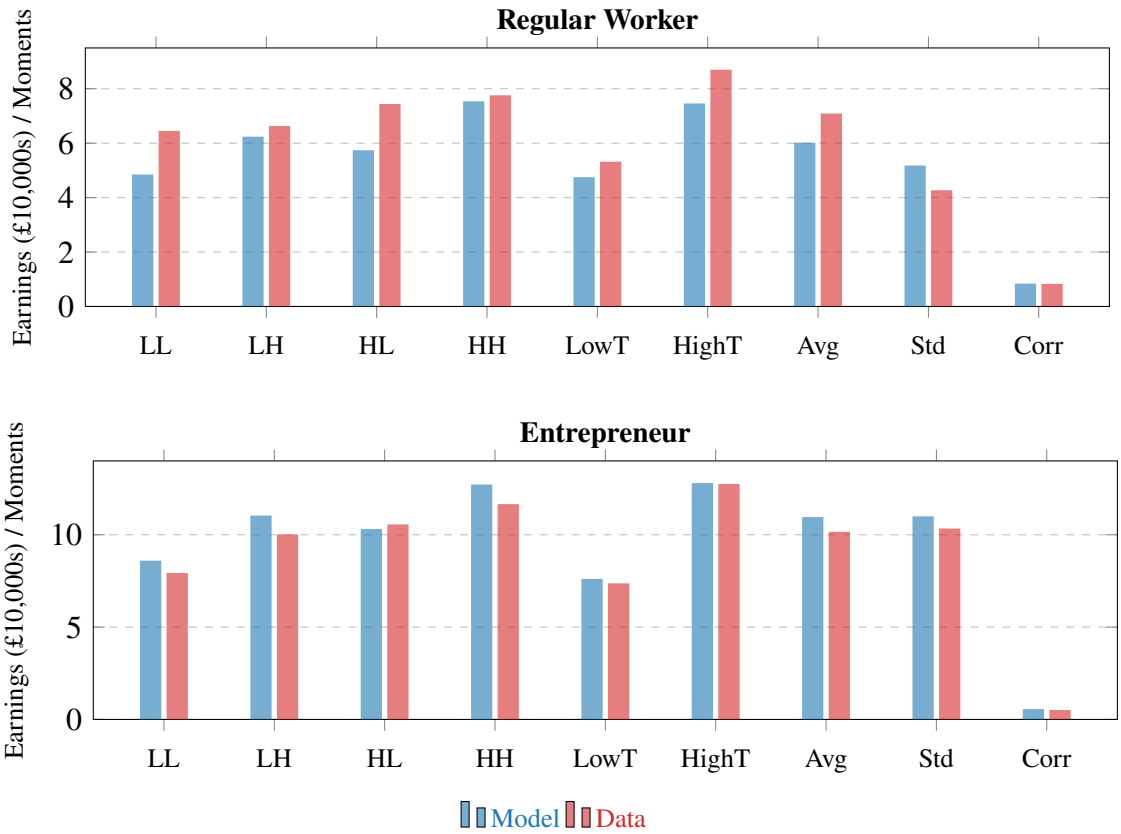
Production function parameters: Firm total productivity (A) and the capital share (α) are not targeted through a separate data moment. Instead, we recover them by inverting the representative firm's two first-order conditions for the wage and rental rate. Given the factor prices (w, r) observed in Understanding Society and the model-implied aggregate capital K_C and labor L_C at the stationary equilibrium, these conditions form a system of two equations in the two unknowns (A, α) that we solve exactly. In effect, A and α are chosen so that the model reproduces observed factor prices at its equilibrium factor quantities.

4.4 Model Fit

Targeted Moments. The model closely matches the empirical moments targeted in estimation. Figure 4 evaluates the fit for earnings profiles across occupations and genetic endowments. The model reproduces both the level and dispersion of earnings for regular workers and entrepreneurs. In particular, it captures the substantial earnings premium enjoyed by entrepreneurs with favorable genetic endowments (High PGS_{edu} and High PGS_{risk}), as well as the pronounced earnings differences between high- and low-type individuals. The model also matches key earnings dynamics. The simulated standard deviation and serial correlation closely align with their empirical counterparts, reproducing the higher volatility and lower persistence of entrepreneurial earnings relative to paid employment.

Figure 5 evaluates the model's ability to reproduce the mechanisms governing intergenerational transmission. Panel A shows that the model closely matches the probability that children become high pre-market types conditional on parental genes and parental type. Panel B demonstrates that the model replicates the observed gradient in parental investments across children's genetic endowments, with children possessing high PGS_{edu} receiving greater parental investment. These patterns are consistent with the model's central mechanism in which parental

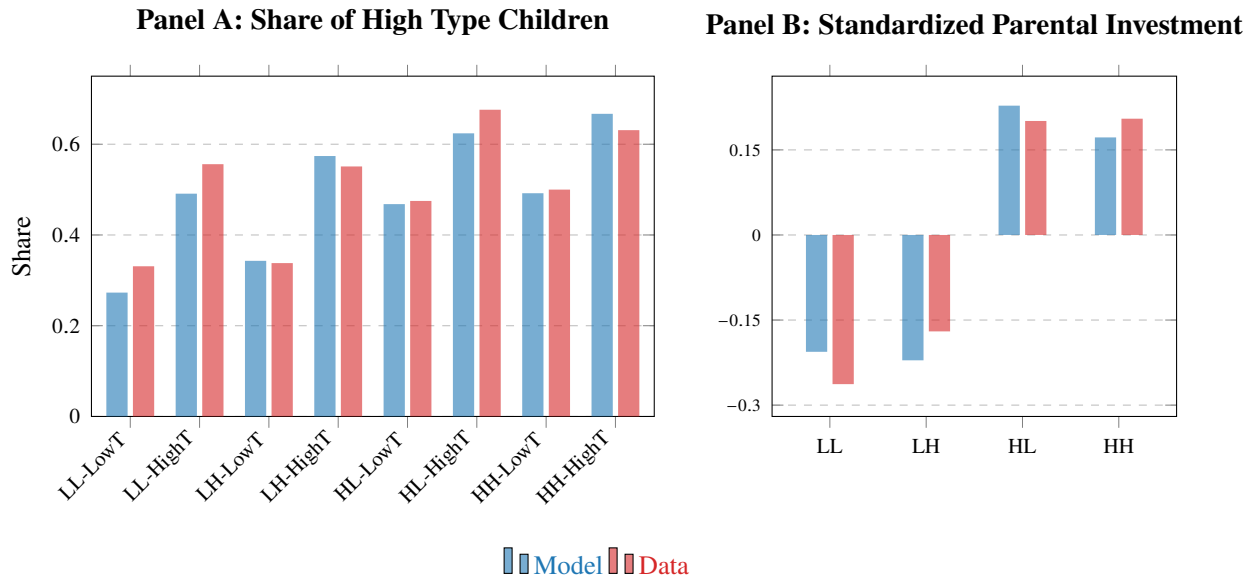
Figure 4: Model Fit: Earnings Profiles



Notes: The figure compares simulated and empirical earnings moments (in 2015 £10,000s) for Regular Workers (top panel) and Entrepreneurs (bottom panel). LL/LH/HL/HH denote genetic endowment combinations (e.g., LL = low PGS_{edu} and low PGS_{risk}). LowT/HighT denote low and high pre-market types. Avg is mean earnings, Std is the standard deviation, and Corr is the serial correlation of earnings.

resources and characteristics interact with child-specific returns to investment to shape human capital accumulation.

Figure 5: Model Fit: Parental Investment and Type Transmission



Notes: The figure compares simulated moments with data targets. Panel A shows the share of children becoming high pre-market types conditional on parent's state. X-axis labels denote parent's genes (PGS_{edu}, PGS_{risk}) and pre-market type (e.g., LL-LowT = Low-Low genes, Low pre-market type). Panel B shows the standardized parental time investment received by children conditional on child's genetic endowment (LL = Low-Low genes, etc.).

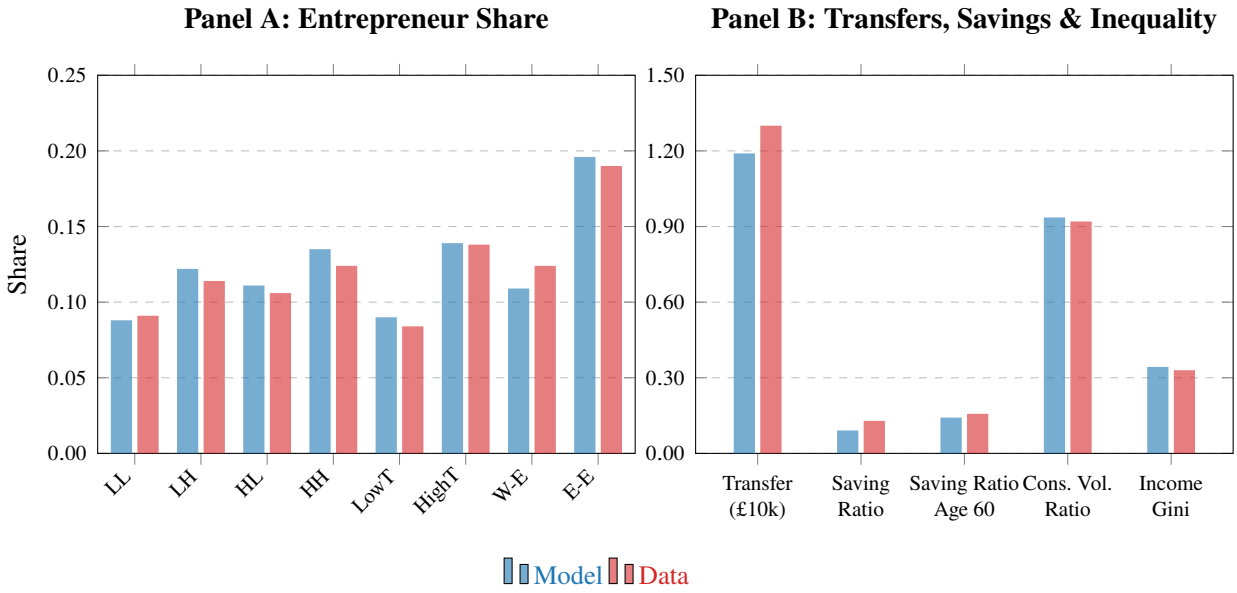
Figure 6 examines occupational sorting, intergenerational persistence, and aggregate household behavior. Panel A shows that individuals with favorable genetic endowments and high pre-market types are substantially more likely to become entrepreneurs. The model also reproduces the strong intergenerational persistence of entrepreneurship observed in the data, with children of entrepreneurs being nearly twice as likely to become entrepreneurs themselves. Panel B shows that the model matches average financial transfers, aggregate saving rates, life-cycle saving profiles, consumption insurance, and income inequality.

The close fit across these moments suggests that the model successfully captures the joint determination of earnings, parental investments, occupational choices, savings behavior, and intergenerational transmission. The mechanisms linking genetic endowments, family investments, and entrepreneurial selection are therefore quantitatively disciplined by the data.

Untargeted Moments. A more demanding test of the framework is whether it can reproduce moments that were not targeted during estimation. We consider two sets of untargeted moments: intergenerational mobility and the wealth distribution. These outcomes are particularly informative because they emerge from the interaction of the model's mechanisms rather than being directly imposed through calibration.

Intergenerational Mobility. Table 6 reports three untargeted measures of intergenerational persistence: the intergenerational elasticity of earnings, the intergenerational elasticity of

Figure 6: Model Fit: Entrepreneurship Shares, Transfers, Savings and Inequality



Notes: Panel A displays the share of entrepreneurs by genetic endowment (LL/LH/HL/HH), pre-market type (LowT/HighT), and intergenerational transition probabilities (W-E = Worker parent to Entrepreneur child; E-E = Entrepreneur to Entrepreneur). Panel B displays the mean financial transfer (in £10,000s), aggregate saving ratio, saving ratio at age 60, the consumption volatility ratio between pre-market types $[\frac{\text{Var}(\log c|\tau=High)}{\text{Var}(\log y|\tau=High)} / \frac{\text{Var}(\log c|\tau=Low)}{\text{Var}(\log y|\tau=Low)}]$, and income Gini. Blue bars represent model simulation; red bars represent data targets.

consumption, and the intergenerational correlation of pre-market types.

The model generates an earnings elasticity of 0.43, remarkably close to the empirical benchmark for the UK.[19] It also closely matches observed persistence in consumption and pre-market types. These moments were not directly targeted during estimation, yet emerge naturally from the interaction of parental investments, occupational choices, savings decisions, and genetic transmission.

Table 6: Model Fit: Untargeted Intergenerational Mobility Moments

Moment	Model	Data
Intergenerational elasticity of earnings (IGE)	0.428	0.430
Intergenerational elasticity of consumption	0.491	0.500
Intergenerational correlation of pre-market type	0.236	0.210

Notes: The table reports model-implied untargeted intergenerational moments alongside empirical benchmarks. The earnings elasticity is based on Gregg et al. (2017). The consumption elasticity benchmark is drawn from Bolt et al. (2024) and Gallipoli et al. (2020). The correlation of pre-market type is estimated using the Understanding Society sample employed in this paper.

Wealth Distribution. The wealth distribution provides an especially demanding untargeted test of the model. Wealth moments are not included among the estimation targets because the Understanding Society survey top-codes wealth and therefore does not reliably capture the upper

[19]The model uses lifetime permanent income, making it more comparable to estimates that reduce transitory shocks and life-cycle bias, such as Dearden et al. (1997) and Gregg et al. (2017).

tail of the distribution. Instead, Table 7 compares model-generated wealth shares with estimates from the World Inequality Database (WID) (Alvaredo et al., 2017), which combines survey, tax, and administrative records to measure wealth concentration more accurately.

Table 7: Top Wealth Shares: Model vs. WID Data

	Model	**Data**
Top wealth shares		
Top 0.1%	0.066	0.065
Top 1%	0.184	0.211
Top 10%	0.495	0.550
Top 20%	0.648	0.700
Top 50%	0.897	0.950
Top 90%	0.998	0.999
Inequality		
Wealth Gini	0.629	0.650

Notes: This table compares the model-generated wealth distribution moments with empirical benchmarks from the World Inequality Database (WID) (Alvaredo et al., 2017). "Top X%" denotes the cumulative share of wealth held by the wealthiest X% of the population.

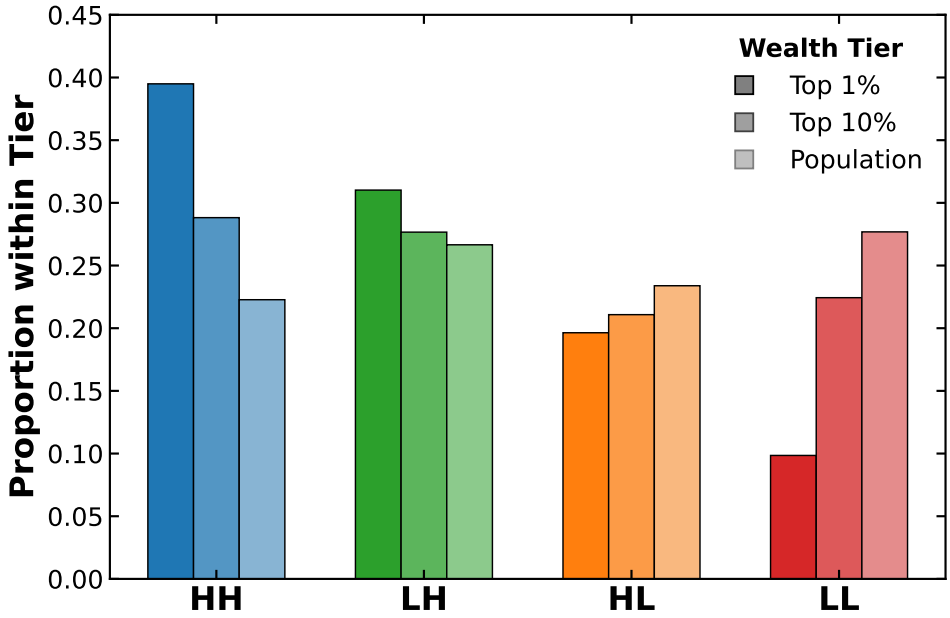
The model closely reproduces top wealth shares and overall wealth inequality. In particular, it generates realistic wealth holdings among the top 0.1%, 1%, and 10% of households, as well as a wealth Gini coefficient close to the empirical benchmark. This agreement is notable because none of these moments are targeted during estimation.

Matching the upper tail is non-trivial. Standard heterogeneous-agent models with thin-tailed shocks typically generate substantially less wealth concentration than observed in the data (Luttmer, 2016; Gabaix et al., 2016). A prominent explanation emphasizes persistent heterogeneity, entrepreneurial risk-taking, and scalable technologies as sources of extreme wealth accumulation (Rosen, 1981; Gabaix and Landier, 2008). Consistent with this view, Fagereng et al. (2020) and Bach et al. (2020) show that a small fraction of households undertakes concentrated idiosyncratic risk and accounts for a disproportionate share of aggregate wealth.

Our framework endogenizes these forces through entrepreneurial selection. Individuals with favorable combinations of educational and risk-tolerance genetic endowments are more likely to enter entrepreneurship, where scalable opportunities amplify persistent heterogeneity and generate non-linear wealth dynamics. Combined with financial frictions, this mechanism produces realistic wealth concentration without relying on exogenous heavy-tailed shocks or directly targeting top wealth moments.

Figure 7 provides additional evidence on the mechanism underlying wealth concentration. High-endowment groups, particularly individuals with high values of both PGS_{edu} and PGS_{risk}, are disproportionately represented among the highest wealth tiers, whereas the LL group becomes progressively less common as wealth increases. This pattern emerges endogenously from the model. Individuals with favorable endowments are more likely to develop high pre-market types, accumulate human capital, and sort into entrepreneurship, where scalable opportunities generate

Figure 7: Genetic Type Composition Across Wealth Tiers



Notes: This figure shows the distribution of genetic types across wealth tiers. Each bar reports the proportion of individuals within a given wealth tier belonging to one of four genetic types defined by polygenic scores for educational attainment (PGS_{edu}) and risk tolerance (PGS_{risk}): HH (high–high), HL (high–low), LH (low–high), and LL (low–low). The vertical axis reports the share of each genetic type within each tier.

larger wealth gains.

Importantly, the figure should not be interpreted as implying a deterministic relationship between genetic endowments and wealth. Most individuals with favorable endowments do not become extremely wealthy, and many wealthy individuals do not belong to the highest-endowment group. Instead, the figure illustrates how genetic endowments shift the probability of entering high-return entrepreneurial trajectories. The over-representation of high-endowment types at the top of the wealth distribution therefore provides additional support for the model's mechanism linking genetic heterogeneity, entrepreneurial selection, and long-run wealth accumulation.

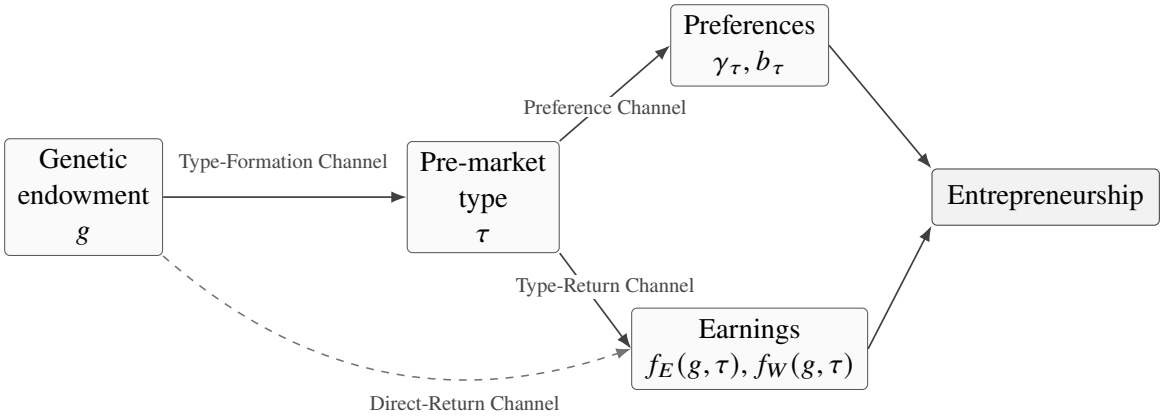
The model also reproduces realistic levels of intergenerational mobility, despite not targeting mobility moments during estimation. The ability to match both wealth concentration and intergenerational persistence is particularly informative because these are two central features of economic inequality. Their emergence from the model provides strong validation of the mechanisms linking genetic endowments, parental investments, occupational choice, and the transmission of economic status across generations.

5 Key Estimation Results

The structural estimates reveal four channels through which genetic endowments affect entrepreneurial choice. Genetic endowments influence the formation of pre-market types, which summarize educational attainment and preferences prior to labor-market entry. These types then affect entrepreneurship through preferences and occupation-specific returns. Genetic

endowments also affect earnings directly, conditional on type. Figure 8 summarizes these mechanisms, which we refer to as the type-formation, preference, type-return, and direct-return channels.[20]

Figure 8: Genetic Heterogeneity and Entrepreneurial Choice



5.1 Type-Formation Channel

The first channel operates through pre-market type formation. Table 8 reports estimates from equation (3.2), where the probability that a child becomes a high pre-market type depends on the child's genetic endowment, parental characteristics, and parental investment.

Three results stand out. First, children's own genetic endowments strongly predict type formation. The largest effects are associated with educational-attainment endowments, with children in the HH and HL groups substantially more likely to become high types than children in the baseline LL group. Second, family background remains important even after conditioning on the child's own genetic endowment. Children of high-type parents are significantly more likely to become high types themselves, and parental genetic endowments generate an additional gradient. These findings are consistent with the role of family background and genetic nurture documented in recent work (Bolt et al., 2025; Houmark et al., 2024). Third, parental investment exhibits strong complementarity with child endowments. Although investment raises the probability of high-type formation for all children, its effect is considerably larger among those with more favorable genetic endowments.

The estimates therefore suggest that pre-market heterogeneity emerges from the interaction of genetic endowments, family background, and parental investments. Rather than operating independently, these factors reinforce one another in shaping the skills and preferences individuals bring into the labor market.

[20]Throughout this section, HH, HL, LH, and LL denote high/low combinations of PGS_{edu} and PGS_{risk}. Coefficients are normalized relative to the low pre-market type and the LL genetic group unless otherwise stated.

Table 8: Estimated Type-Formation Channel

	Estimate	Std. Err.
Child genetic endowment $[\delta_{g^c}]$		
Child HH	0.390	(0.039)
Child HL	0.379	(0.053)
Child LH	0.086	(0.032)
Parental background $[\delta_\tau,\ \delta_g]$		
High parental type	1.258	(0.035)
Parent HH	0.689	(0.042)
Parent HL	0.545	(0.027)
Parent LH	0.356	(0.056)
Parental investment and complementarity $[\delta^{t^p}_{g^c}]$		
Investment × Child HH	12.387	(1.009)
Investment × Child HL	2.810	(0.109)
Investment × Child LH	4.127	(1.215)
Investment × Child LL	1.473	(0.029)

Notes: This table reports normalized estimates from equation (3.2). The outcome is high pre-market type classification. The omitted category is a child in the LL genetic group with a low-type parent in the parental LL genetic group. The investment coefficients report effects on the log-odds of high pre-market type formation.

5.2 Preference Channel

Pre-market types also influence entrepreneurship through preferences. In the model, γ_τ governs risk preference and b_τ captures the non-pecuniary value of entrepreneurship in equation (3.1). These parameters are important because entrepreneurship exposes households to substantial income risk while also providing non-pecuniary benefits or costs that are not reflected in earnings alone.

Table 9 reports the estimated preference parameters. High pre-market types are significantly less risk averse than low pre-market types, making entrepreneurial income risk less costly to bear. The non-pecuniary estimates reinforce the same sorting force. High types derive positive utility from entrepreneurship, whereas low types experience a negative non-pecuniary value. At median consumption, these estimates imply that the non-pecuniary value of entrepreneurship is equivalent to a 4.4 percent consumption gain for high types and a 13.7 percent consumption penalty for low types; Appendix Figure D.2 reports the full consumption-equivalent calculation.

These estimates imply that high types are both more willing to bear entrepreneurial risk and more likely to value entrepreneurship itself, generating an important non-pecuniary source of entrepreneurial selection.

Table 9: Estimated Preference Channel

	Estimate	Std. Err.
Risk aversion $[\gamma_\tau]$		
γ_H	2.972	(0.015)
γ_L	3.446	(0.028)
$\gamma_H - \gamma_L$	-0.475	(0.018)
Non-pecuniary value of entrepreneurship $[b_\tau]$		
b_H	9.58×10^{-4}	(9.9×10^{-5})
b_L	-4.62×10^{-3}	(5.87×10^{-4})
$b_H - b_L$	5.58×10^{-3}	(5.95×10^{-4})

Notes: This table reports estimated preference parameters from equation (3.1).

5.3 Earnings Return Channels

The final set of estimates quantifies how genetic endowments and pre-market types shape earnings in entrepreneurship and wage work. Equation (3.4) allows both g and τ to shift occupation-specific efficiency units. This separates the type-return channel, operating through $\beta_{\tau,\omega}$, from the direct-return channel, operating through $\beta_{g,\omega}$.

Table 10 highlights three findings. First, genetic endowments directly affect earnings in both sectors. The largest premia arise for the HH and LH groups, particularly in entrepreneurship, suggesting that risk-related endowments are especially valuable in entrepreneurial activities. Second, high pre-market types earn more in both sectors, although the premium is larger in wage work. This pattern suggests that the human-capital component captured by pre-market type is more strongly rewarded in paid employment, whereas genetic endowments contribute relatively more to comparative advantage in entrepreneurship.

Third, the earnings processes differ sharply across occupations. Wage earnings are highly persistent and relatively stable over the life cycle, whereas entrepreneurial earnings are substantially more volatile and less persistent. Moreover, worker-specific productivity shocks exhibit almost no intergenerational persistence once pre-market human capital is accounted for, consistent with Bolt et al. (2024). In contrast, entrepreneurial productivity shocks remain strongly persistent across generations. This suggests the existence of an additional dynastic component of entrepreneurial success beyond measured genetic endowments, pre-market types, and wealth. Such persistence may reflect the transmission of business-specific knowledge, managerial practices, customer and supplier networks, or business opportunities across generations.

The estimates imply that entrepreneurial selection operates through both preferences and comparative advantage. High pre-market types are more willing to bear entrepreneurial risk and derive greater utility from entrepreneurship, while favorable genetic endowments increase entrepreneurial returns conditional on type. Entrepreneurship therefore emerges from the interaction of preferences, skills, and earnings opportunities rather than from any single channel

alone.

Table 10: Estimated Return Channels

	Entrepreneurship		Wage Work	
	Estimate	Std. Err.	Estimate	Std. Err.
Direct genetic returns $[\beta_{g,\omega}]$				
HH	0.381	(0.026)	0.367	(0.020)
HL	0.169	(0.027)	0.109	(0.020)
LH	0.342	(0.026)	0.296	(0.023)
Type returns $[\beta_{\tau,\omega}]$				
High pre-market type	0.431	(0.022)	0.538	(0.006)
Sector dynamics				
Within-life persistence $[\rho_\omega]$	0.355	(0.013)	0.919	(0.005)
Innovation std. dev. $[\sigma_\omega]$	0.607	(0.009)	0.452	(0.003)
Intergenerational persistence $[\rho^c_\omega]$	0.859	(0.031)	0.018	(0.001)
Return to entrepreneurial capital $[\nu]$	0.624	(0.015)	–	

Notes: This table reports normalized estimates from the occupation-specific earnings processes in equation (3.4). The omitted category is the LL genetic group and the low pre-market type.

6 The Impact of Genetic Heterogeneity in Explaining Mobility and Wealth Concentration

The previous section established that genetic endowments affect economic outcomes through three interconnected channels: pre-market type formation, entrepreneurial selection and persistence, and occupation-specific earnings. We now quantify the aggregate importance of these mechanisms. To do so, we compare the benchmark economy to two counterfactual economies that sequentially eliminate (i) the intergenerational transmission of genetic endowments and (ii) individual genetic heterogeneity itself.

Our decomposition isolates two conceptually distinct forces. The first is the *intergenerational genetic link*, which creates persistence in endowments across generations and generates dynastic sorting. The second is *individual genetic heterogeneity*, which creates differences in skills, preferences, and productivity within each generation. By comparing the benchmark economy to these counterfactuals, we quantify how much each force contributes to intergenerational mobility, individual life-cycle outcomes, and wealth concentration.

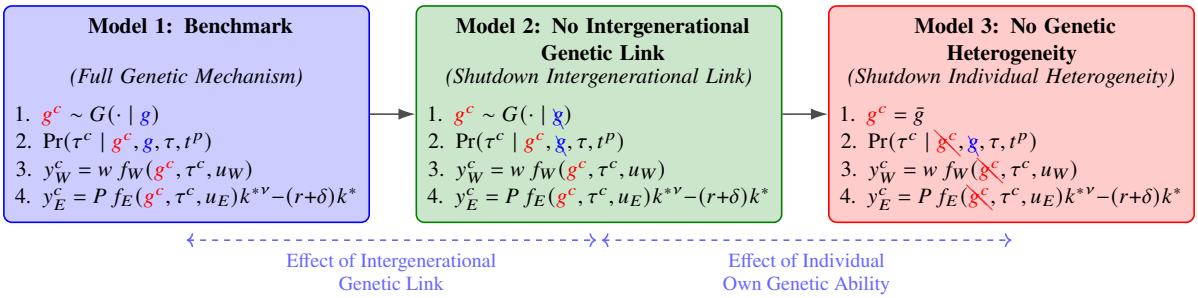
6.1 Decomposition Strategy

We implement the sequential decomposition illustrated in Figure 9. Model 2 removes the intergenerational genetic link while preserving individual genetic heterogeneity. Children

continue to differ genetically, but their endowments are no longer systematically related to those of their parents. Model 3 additionally removes individual genetic heterogeneity by assigning all individuals the population-average genetic endowment and eliminating genetic inputs from the type-formation and earnings functions.

Each counterfactual economy is solved to its own steady state. As a result, pre-market types, occupational choices, earnings, and wealth distributions adjust endogenously to the modified genetic structure. The decomposition therefore allows us to distinguish between the long-run dynastic effects of genetic transmission and the within-generation effects of cross-sectional genetic heterogeneity.

Figure 9: Decomposition Strategy: Isolating Genetic Channels



Notes: Model 1 is the benchmark economy. Model 2 removes the intergenerational genetic link by replacing the conditional inheritance distribution $G(\cdot \mid g)$ with the unconditional population distribution and dropping parental genes from the type-formation function; individual genetic heterogeneity is preserved, but endowments are no longer transmitted across generations. Model 3 additionally removes individual genetic heterogeneity by assigning every agent the population-mean endowment $g^c = \bar{g}$ and eliminating genetic inputs from the type-formation and earnings functions. Each counterfactual is solved to a new steady state, so that pre-market types τ and initial financial transfers a_1 adjust endogenously to the modified genetic structure.

6.2 Impact on Intergenerational Mobility

We begin by examining intergenerational mobility. Table 11 shows that genetic mechanisms contribute substantially to persistence in economic status across generations.

Panel A focuses on earnings mobility. Shutting down both genetic channels relative to the benchmark (Δ_{13} in Table 11) reduces the intergenerational elasticities of earnings and consumption by roughly one quarter and one fifth, respectively, and lowers the rank–rank correlation by nearly one third.[21] At the same time, upward mobility rises by more than 8%. These results indicate that genetic mechanisms increase persistence in economic status and reduce the likelihood that children born to lower-income families move up the earnings distribution.

Panel B shows that genetic mechanisms also affect entrepreneurship-specific mobility, which we summarize with two intergenerational transition probabilities: the probability that a child becomes a *long-duration entrepreneur*—one who runs a business for more than half

[21] Following Chetty et al. (2014), we rank parents and children by earnings within their respective generation-specific distributions. The rank–rank correlation is the Pearson correlation between parent and child percentile ranks. Upward mobility is the predicted child percentile rank for a parent at the 25th percentile of the parent rank distribution.

of the working life—conditional on a long-duration entrepreneur parent ($P(E \to E)$) or on a worker parent ($P(W \to E)$). Removing genetic heterogeneity reduces $P(E \to E)$, weakening entrepreneurial persistence across generations. Conversely, removing the intergenerational genetic link increases $P(W \to E)$, raising transitions from worker families into entrepreneurship. In the benchmark economy, favorable entrepreneurial endowments tend to remain concentrated within entrepreneurial families through inheritance and assortative matching. Once this dynastic link is removed, children from worker and entrepreneur families draw endowments from the same population distribution, increasing entrepreneurial mobility.

Table 11: Intergenerational Mobility: Role of Genetic Heterogeneity

Outcome	Model 1 Benchmark	Model 2 No IG Link	Δ_{12}	Model 3 No G Heterogeneity	Δ_{23}	Δ_{13}
Panel A: Earnings Mobility						
IGE	0.427	0.397	−7.1%	0.319	−18.1%	−25.3%
IGC	0.490	0.463	−5.6%	0.398	−13.3%	−18.9%
Rank–rank correlation	0.407	0.366	−10.1%	0.284	−20.1%	−30.2%
Upward mobility	0.397	0.408	+2.8%	0.430	+5.4%	+8.2%
Panel B: Entrepreneurship Mobility						
$P(E \to E)$	0.0245	0.0241	−1.6%	0.0214	−11.0%	−12.7%
$P(W \to E)$	0.0076	0.0084	+10.5%	0.0082	−2.6%	+7.9%
Panel C: Pre-Market Type Transmission						
Type correlation	0.247	0.196	−20.6%	0.156	−16.2%	−36.8%

Notes: Model 1 is the benchmark; Model 2 removes the intergenerational genetic link; Model 3 additionally removes individual genetic heterogeneity (see Figure 9). The columns Δ_{12}, Δ_{23}, and Δ_{13} report percentage changes relative to Model 1 for the corresponding model pairs, so negative values indicate that the outcome falls as genetic channels are removed. IGE and IGC are the elasticities of child log earnings and log consumption with respect to the parent's; the rank–rank correlation uses within-cohort earnings ranks; and upward mobility is the predicted percentile rank of a child whose parent ranked at the 25th percentile of the parent earnings distribution. Type correlation is the parent–child correlation in high/low pre-market type. $P(E \to E)$ and $P(W \to E)$ are the probabilities that a child is a long-duration entrepreneur—an entrepreneur for more than half of the working life—conditional on an entrepreneur or worker parent, respectively.

Panel C highlights one of the key mechanisms underlying these patterns. The parent–child correlation in pre-market type falls by more than one third when genetic heterogeneity is removed. This decline reflects both the disappearance of dynastic transmission and the elimination of genetic differences that shape skill formation within generations. Genetic endowments therefore influence mobility well before labor-market entry through their effects on education- and risk-related traits.

Taken together, the results suggest that both dynastic transmission and individual genetic heterogeneity contribute to intergenerational persistence. Comparing the two decomposition steps shows which channel dominates, and the answer differs by outcome. For earnings and consumption mobility, individual heterogeneity (Δ_{23}) accounts for the larger share of the total genetic contribution, while the dynastic link (Δ_{12}) plays the smaller role. The ordering reverses for pre-market type transmission, where severing the dynastic link is the more important step. This contrast is intuitive: because genes pass directly from parent to child, removing the dynastic link bears most heavily on the parent–child correlation in endowments themselves, whereas

individual heterogeneity matters most for the dispersion of realized earnings and consumption. Overall, genetic mechanisms matter not only because favorable endowments are transmitted across generations, but also because they create persistent differences in skills, preferences, and entrepreneurial success within each generation.

6.3 Impact on Individual Lifetime Outcomes

The previous subsection showed that genetic heterogeneity contributes to intergenerational persistence. We now examine the underlying life-cycle mechanisms by quantifying how much variation in individual outcomes is attributable to genetic channels. Structural life-cycle models typically rely on latent, unobserved heterogeneity to account for dispersion in long-run outcomes such as earnings, wealth, and utility (Huggett et al., 2011; Keane and Wolpin, 1997; Todd and Zhang, 2020). An important question is therefore how much of this heterogeneity can be linked to measured genetic endowments.

We isolate the contribution of genetic heterogeneity by comparing the benchmark (Model 1) with the specification that removes individual genetic heterogeneity (Model 3). Table 12 reports the results. Genetic heterogeneity explains 6.7% of the variation in pre-market type, indicating that genetic endowments play an important role in shaping skills and preferences before labor-market entry. This finding suggests that part of the heterogeneity commonly treated as unobserved in structural models originates from differences that are already present early in life and persist throughout the life cycle.

The effects become even more pronounced when we examine lifetime economic outcomes. Genetic heterogeneity explains 8.6% of the variation in lifetime earnings in the full population, compared with 6.9% among wage workers and 14.1% among entrepreneurs.[22] This pattern is consistent with the return-channel estimates in Section 5.3. Genetic endowments affect earnings in both occupations, but entrepreneurial careers amplify these differences because productive advantages can be scaled through capital accumulation and because entrepreneurial returns are more sensitive to heterogeneity in risk-related traits. Consequently, entrepreneurship acts as a mechanism through which relatively modest differences in endowments are transformed into much larger differences in lifetime economic outcomes.

Finally, genetic heterogeneity explains 5.5% of the variation in lifetime utility. Although smaller than its contribution to earnings, this effect remains economically meaningful because lifetime utility incorporates consumption smoothing, occupational choices, risk exposure, and non-pecuniary returns. The result suggests that measured genetic endowments account for a non-trivial share of the latent heterogeneity that structural life-cycle models rely on to explain dispersion in long-run outcomes.

It is important to interpret these magnitudes relative to the latent types typically used in structural life-cycle models. First, we consider only two measures of genetic endowments, PGS_{edu} and PGS_{risk}. Since these scores capture only a subset of the genetic variation relevant to

[22] A useful benchmark is Bolt et al. (2025), who find that genetic endowments explain approximately 5% of lifetime earnings variation among wage workers.

economic behavior, the full contribution of genetic endowments is likely larger. Morevoer, in our framework, genetic endowments are measured at birth and therefore capture only the component of heterogeneity present before subsequent investments and experiences are realized. By contrast, the observable or latent types used in many structural models are effectively measured much later in life, often at labor-market entry, after years of parental investments, schooling, and other environmental influences have already shaped skills and preferences. Consequently, it is natural that measured genetic endowments explain only a subset of the heterogeneity captured by these later-life types. The remaining gap reflects the cumulative effects of environmental influences and gene–environment interactions that unfold throughout childhood and adolescence.

Taken together, the results reveal a clear amplification mechanism. Genetic heterogeneity affects pre-market skill formation, influences productivity throughout the working life, and generates particularly large differences among entrepreneurs. These life-cycle effects provide the micro-foundations for the mobility and wealth-concentration results documented elsewhere in the paper. In particular, the disproportionately large contribution of genetic heterogeneity to entrepreneurial earnings points to entrepreneurship as a key channel through which genetic differences translate into persistent wealth inequality.

Table 12: Variance Explained by Genetic Heterogeneity

Outcome	Variance explained (%)
Pre-market type	6.73
Lifetime earnings, all	8.59
Wage workers	6.91
Entrepreneurs	14.12
Lifetime utility	5.45

Notes: This table reports the share of cross-sectional variation in life-cycle outcomes associated with genetic heterogeneity. Lifetime earnings are the present discounted value of labor income, discounted at the market interest rate r. Lifetime utility is the present discounted value of flow utility, discounted at β.

6.4 Impact on Wealth Concentration

The previous subsection showed that genetic heterogeneity has its largest effects on lifetime earnings among entrepreneurs, where productive advantages can be scaled through capital accumulation. We now examine whether these amplified earnings differences translate into differences in wealth accumulation and, ultimately, wealth concentration.

The model predicts that genetic heterogeneity should matter disproportionately at the top of the wealth distribution. For most households, wealth accumulation occurs primarily through labor income and precautionary saving, leaving limited scope for small productivity differences to generate large wealth gaps. Entrepreneurial households, however, can scale productive opportunities through capital accumulation. Persistent differences in entrepreneurial productivity therefore compound over time and across generations, implying that genetic mechanisms should be most visible among the wealthiest households.

Table 13 confirms this prediction. In the benchmark economy, wealth is highly concentrated, with the top 1% holding 18.4% of total wealth. Removing the intergenerational genetic link lowers the top 1% share to 17.0%, while removing genetic heterogeneity entirely reduces it to 13.7%. Below the top 1%, the effects are modest. The combined top 1–20% share changes little, and the increases in the lower wealth bins mainly reflect a mechanical reallocation of wealth shares as the extreme upper tail contracts.

The two channels operate differently. Removing individual genetic heterogeneity (Model 3) lowers top wealth shares more than removing the intergenerational genetic link alone (Model 2), so within-generation differences in endowments account for most of the reduction in top-share concentration. Both channels nonetheless contribute, mirroring the mobility findings from the previous subsection: genetic endowments matter not only because they create differences within a generation, but also because favorable entrepreneurial traits are transmitted across generations through genetic transmission, parental investment, and financial transfers, allowing advantages to persist and compound over time.

Table 13: Wealth Distribution Across Genetic Counterfactuals

	Model 1 Benchmark	Model 2 No IG Link	Model 3 No G Heterogeneity	Difference M1 − M3
Wealth shares				
Top 1%	0.184	0.170	0.137	0.047
Top 1–10%	0.311	0.314	0.309	0.002
Top 10–20%	0.153	0.157	0.145	0.008
Top 20–50%	0.249	0.255	0.282	-0.033
Top 50–90%	0.101	0.102	0.124	-0.023
Bottom 10%	0.002	0.002	0.003	-0.001
Inequality				
Wealth Gini	0.629	0.622	0.576	0.053

Notes: This table reports model-generated wealth distribution moments across the benchmark economy and two genetic counterfactuals. Wealth shares are reported for mutually exclusive wealth quantile bins. "No IG Link" removes the intergenerational genetic link, so child genetic endowments are drawn independently of parental genetic endowments. "No G Heterogeneity" removes individual genetic heterogeneity. The final column reports the difference between the benchmark model and the no genetic heterogeneity counterfactual.

The effects become substantially smaller as we move down the wealth distribution. Appendix E.1 reports two further decompositions that reinforce this pattern. Mean wealth ratios show that the top 0.1% holds roughly 118 times the average wealth of the bottom 90% in the benchmark economy, falling to 89 once the intergenerational genetic link is removed and to 45 once all genetic heterogeneity is eliminated. A variance decomposition shows that the genetic contribution to wealth dispersion is mostly concentrated within the top 1%, with only a modest contribution from the bottom 99%. This also indicates that the decline in the wealth Gini, from 0.629 to 0.576, is primarily driven by reduced wealth dispersion at the top of the distribution. Taken together, these results show that genetic mechanisms matter for wealth concentration, but their quantitative importance is concentrated in the upper tail rather than spread evenly across

the population.[23]

These findings connect naturally to theories emphasizing scale, talent, and superstar effects. Models beginning with Rosen (1981) show how small differences in ability can generate disproportionately large rewards when opportunities are scalable. Empirical work on entrepreneurs, executives, and business owners similarly documents how expanding market scale amplifies rewards at the top (e.g., Gabaix and Landier, 2008; Kaplan and Rauh, 2013). Our framework adds a dynastic dimension to this literature. Persistent entrepreneurial advantages, partly linked to measured genetic endowments, generate differences that compound through both capital accumulation and intergenerational transmission, helping to sustain wealth concentration at the top of the distribution.

7 Policy Experiments: Financing Wealth versus Financing Talent

The previous sections showed that entrepreneurial ability is an important source of mobility, earnings differences, and wealth concentration. These findings raise a natural policy question: when entrepreneurial talent and wealth are imperfectly aligned, how should capital be allocated? Should financing be tied primarily to collateral and existing wealth, or should it be directed toward entrepreneurs with the highest expected returns?

We address this question by comparing two financing institutions that embody opposite allocation rules. A *wealth-based* rule ties access to capital to collateral and existing assets; we implement it through a broad relaxation of the collateral constraint. A *talent-based* rule ties access to expected entrepreneurial returns; we implement it through a venture-capital sector that supplies capital in exchange for a share of the business. The section therefore studies how the *allocation* of entrepreneurial capital, rather than its quantity, shapes mobility, inequality, and aggregate output.

Financing Wealth versus Financing Talent. The wealth-based institution is a broad credit relaxation. Starting from the benchmark credit constraint $\chi^{bm} = 0.22$, it raises the borrowing limit uniformly:

$$k_j \leq (1+\chi)a_j, \qquad \chi \geq \chi^{bm}.$$

Because the constraint scales with assets, this policy expands financing capacity through collateral, so access remains tied to wealth.

The talent-based institution is a venture-capital channel. The benchmark collateral constraint remains fixed at

$$\bar{k}^{bm}(a_j) = (1+\chi^{bm})a_j,$$

[23]The variance decomposition further shows that, while individual genetic heterogeneity drives more of the reduction in top wealth *shares*, the intergenerational genetic link accounts for most of the genetic contribution to wealth *variance*, reflecting its outsized role in generating the most extreme fortunes; see Appendix E.1.

but an entrepreneur who wishes to operate beyond what collateral can finance can raise the shortfall

$$\ell_j = \left[k_j - \bar{k}^{bm}(a_j)\right]_+$$

from a VC sector. The VC investor provides money in exchange for an equity share s_j of the business. For a given entrepreneur and capital choice k_j, a competitive VC sector offers the minimum equity share that breaks even:

$$s_j P f_{E,j} k_j^{\nu} = (r + \delta + r_p)\ell_j,$$

where $r_p \geq 0$ is the excess return required by VC investors. Hence, when $\ell_j > 0$, the required equity share is

$$s_j(k_j; r_p) = \frac{(r + \delta + r_p)\left[k_j - \bar{k}^{bm}(a_j)\right]}{P f_{E,j} k_j^{\nu}}.$$

If $s_j(k_j; r_p) > 1$, the VC contract is infeasible. Thus, VC access is not determined by collateral alone; it also depends on whether the entrepreneurial ability is high enough to make outside equity finance viable.

Capital Allocation. Before turning to mobility and output, we first verify that the two policies differ in the way they allocate capital. We use the entrepreneurial ability specification from equation (3.4): for log entrepreneurial ability,

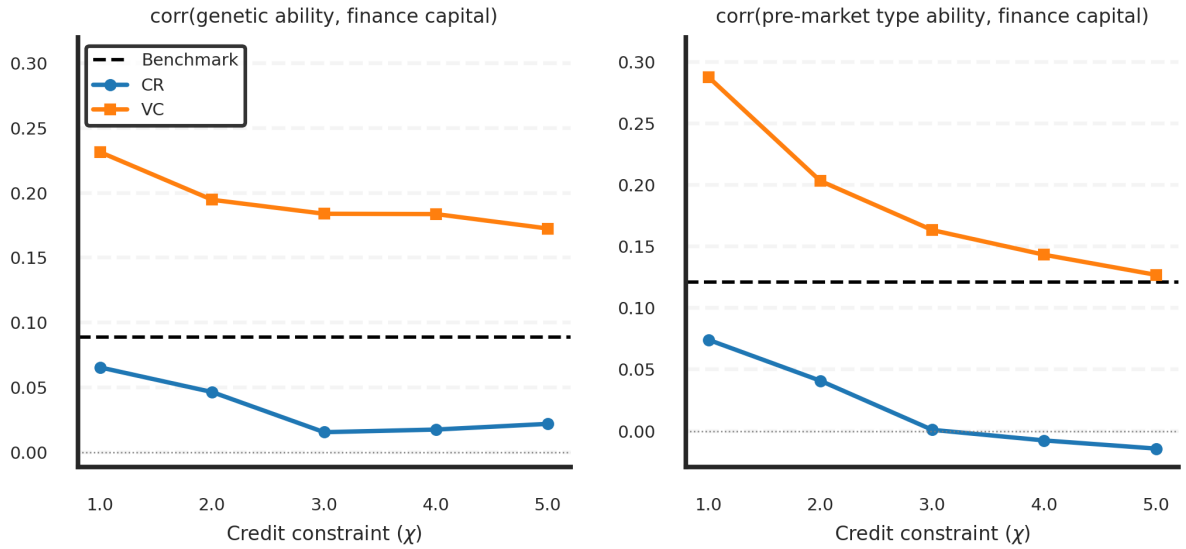
$$\ln f_{E,j} = \beta_{0,E} + \underbrace{\beta_{g,E}}_{\text{return to genetic endowment}} + \underbrace{\beta_{\tau,E}}_{\text{return to pre-market type}} + \underbrace{u_{E,j}}_{\text{shock}}. \tag{7.1}$$

Let $q_j \equiv [k_j - a_j]_+$ denote external entrepreneurial finance. The first two terms, $\beta_{g,E}$ and $\beta_{\tau,E}$, capture the persistent components of entrepreneurial ability associated with genetic endowments and pre-market type. If a financing regime allocates capital toward talent, external finance should be more strongly correlated with these ability components.

Figure 10 shows that this holds only under VC. VC finance generates a much stronger correlation between external finance and both the genetic and the pre-market-type components of entrepreneurial ability than broad credit relaxation. The reason is the difference in allocation rules: credit relaxation expands access through assets, so capital follows wealth, whereas VC finance expands access through expected business returns, so capital follows talent. VC therefore acts as a screening device for entrepreneurial ability even though it does not observe genetic endowments directly.

Decomposing this correlation into its genetic, pre-market-type, and transitory-shock contributions sharpens the contrast. Because equation (7.1) is additively separable, the overall ability–finance correlation is exactly the sum of the three component-specific contributions. Under credit relaxation it is increasingly driven by transitory shocks, whereas under VC it continues to load on the persistent genetic and pre-market-type components. Credit relaxation

Figure 10: Correlation Between Ability Components and External Entrepreneurial Finance



Notes: The figure reports correlations between external entrepreneurial finance $q_j = [k_j - a_j]_+$ and two persistent components of entrepreneurial ability. The left panel reports the correlation between q_j and the genetic component $\beta_{g,E}$ in entrepreneurial ability. The right panel reports the correlation between q_j and the pre-market type component $\beta_{\tau,E}$ in entrepreneurial ability. The dashed black line denotes the benchmark economy. The blue line reports broad credit relaxation (CR). The orange line reports venture-capital financing (VC). For each value of χ, the VC economy is matched to the CR economy by aggregate external entrepreneurial financing.

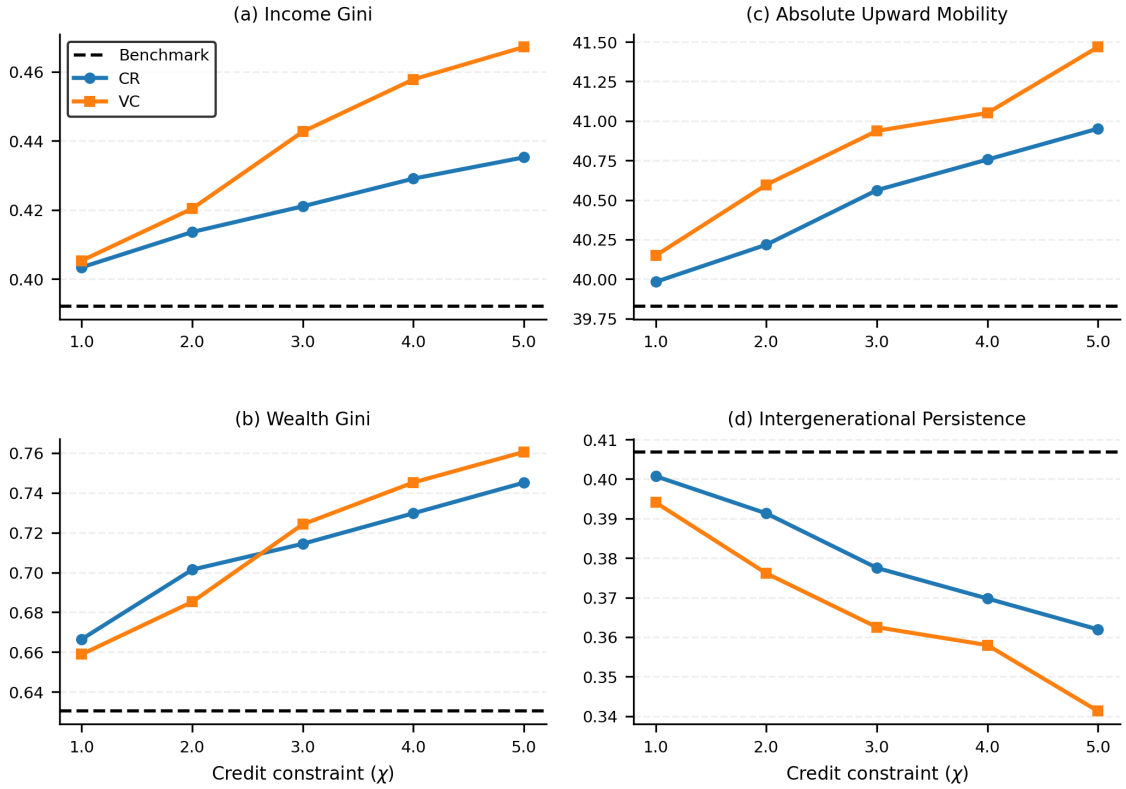
thus makes capital follow wealth and transitory luck, whereas VC makes capital follow durable entrepreneurial talent; the effects on mobility, inequality, and output documented below all stem from this difference. Appendix E.3 reports the full decomposition (Figure E.2) and its derivation.

Mobility and Inequality. Having established that VC allocates capital toward talent, we now trace the consequences for mobility and inequality. Following Chetty et al. (2014), we measure absolute upward mobility as the predicted income rank of a child whose parents are at the 25th percentile of the parent income distribution, and intergenerational persistence as the rank–rank slope between parent and child income ranks. Figure 11 reports these mobility measures together with the income and wealth Gini coefficients across the two financing regimes.

Both financing instruments improve mobility relative to the benchmark, but VC does so more strongly: for the same aggregate financing, it generates higher upward mobility and lower persistence than broad credit relaxation. This follows directly from the allocation result above. Because VC ties capital to talent rather than to wealth, it weakens the link between parental resources and children's entrepreneurial outcomes.

The mobility gain comes with a distributional cost. Panels (a) and (b) show that both policies raise income and wealth inequality, and the increase is substantially larger under VC. By concentrating capital and returns among high-ability entrepreneurs, the same mechanism that improves mobility also widens cross-sectional inequality.

Figure 11: Financing Policy, Mobility, and Inequality

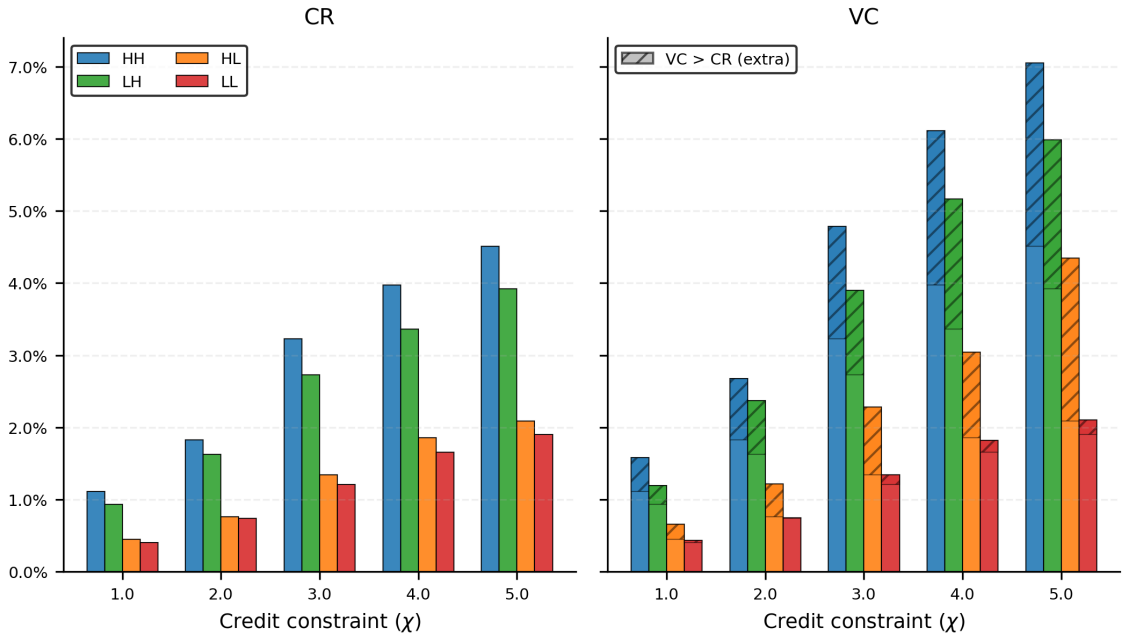


Notes: The figure compares broad credit relaxation (CR) and venture-capital financing (VC). The dashed black line denotes the UK-calibrated benchmark economy. The blue line reports the credit-relaxation counterfactual, which raises the credit constraint χ. The orange line reports the VC counterfactual, where VC investors provide money in exchange for a share of the business. For each value of χ, the VC premium is chosen so that the VC economy delivers the same aggregate amount of external entrepreneurial financing as the corresponding credit-relaxation economy. In the VC panel, the premium revenue collected from entrepreneurs is redistributed to worker households as a dividend proportional to worker household's wealth. Panel (a) reports the income Gini. Panel (b) reports the wealth Gini. Panel (c) reports absolute upward mobility, defined following Chetty et al. (2014) as the predicted income rank of a child whose parents are at the 25th percentile of the parent income distribution. Panel (d) reports intergenerational persistence, measured by the rank–rank slope between parent and child income ranks.

Output and Growth. The talent-based allocation also raises aggregate output. Figure 12 shows aggregate output growth by genetic group. For the same total amount of external entrepreneurial financing, VC generates higher output growth than broad credit relaxation at every financing level. The additional growth is concentrated among individuals with favorable genetic endowments, especially those with high PGS_{edu}, high PGS_{risk}, or both. By contrast, the LL group contributes little to the extra output growth and receives less output growth under VC than under broad credit relaxation. Thus, the aggregate output advantage of VC comes from reallocating capital toward groups with higher entrepreneurial returns.

Discussion: The Mobility–Efficiency–Inequality Tradeoff. The comparison reveals a three-way tradeoff. Financing talent rather than wealth improves intergenerational mobility and aggregate efficiency, because capital is allocated toward entrepreneurs with higher expected returns. However, the same mechanism increases inequality, because the gains accrue dis-

Figure 12: Aggregate Output Growth by Genetic Group



Notes: The figure reports aggregate output growth relative to the benchmark economy, decomposed by genetic group. The left panel reports broad credit relaxation (CR). The right panel reports venture-capital financing (VC), where VC investors provide money in exchange for a share of the business. For each value of the collateral parameter χ, the VC premium is chosen so that the VC economy delivers the same aggregate amount of external entrepreneurial financing as the corresponding CR economy. In the VC panel, the premium revenue collected from entrepreneurs is redistributed to worker households as a dividend proportional to worker household's wealth. Genetic groups are defined by high and low values of PGS_{edu} and PGS_{risk}: HH, HL, LH, and LL. In the VC panel, diagonal hatching marks the portion of output growth that exceeds the corresponding CR value.

proportionately to high-ability entrepreneurs. Thus, policies that improve the allocation of entrepreneurial capital need not reduce inequality, even when they expand opportunity. This coexistence of high mobility and high inequality connects our findings to a literature showing that inequality and mobility need not be inversely related and can rise together when inequality reflects returns to talent (Aghion et al., 2019; Jones and Kim, 2018; Hassler et al., 2007). Our model gives this logic a concrete form: because financing talent rewards entrepreneurial ability rather than existing wealth, it widens the top of the distribution while opening it to newcomers.

8 Conclusion

This paper studies how genetic endowments shape entrepreneurship, intergenerational mobility, and wealth concentration. We combine polygenic scores for educational attainment and risk tolerance with a structural overlapping-generations model in which individuals differ in genetic endowments, pre-market skill formation, preferences, and wealth. The framework allows us to quantify how measured genetic heterogeneity contributes to entrepreneurial selection and success, and to distinguish the roles of individual heterogeneity and intergenerational transmission in generating lifetime inequality and wealth concentration.

Our analysis delivers three main findings. First, genetic heterogeneity is an important source of economic persistence across generations. Genetic endowments influence entrepreneurial outcomes through multiple channels, including skill formation, preferences, and occupation-specific returns. Removing genetic heterogeneity substantially reduces intergenerational persistence in earnings, consumption, and entrepreneurial status while increasing upward mobility. At the individual level, genetic heterogeneity explains a meaningful share of lifetime variation in earnings and welfare. These effects are particularly pronounced among entrepreneurs, whose returns are more sensitive to underlying differences in skills and risk-taking traits. More broadly, our results suggest that measured genetic variation captures an economically important component of the latent heterogeneity commonly invoked in structural models of inequality and mobility.

Second, genetic heterogeneity plays a central role in generating wealth concentration. While its effects on wealth are modest for most households, they become increasingly important toward the top of the distribution. Eliminating genetic heterogeneity reduces the top 1% wealth share from 18.4 percent to 13.7 percent, with much smaller changes elsewhere in the distribution. Decomposition exercises reveal that both within-generation heterogeneity and intergenerational transmission contribute to wealth inequality, but their importance differs across outcomes. Individual heterogeneity accounts for most of the reduction in top wealth shares. These findings highlight how persistent entrepreneurial advantages can be amplified across generations through both selection and inheritance mechanisms.

Third, the allocation of capital has important consequences for mobility, inequality, and economic efficiency. We compare two policies that provide the same aggregate amount of entrepreneurial financing but differ in how capital is allocated. Broad credit relaxation allocates financing primarily on the basis of collateral and existing wealth, whereas venture-capital financing allocates capital based on expected entrepreneurial returns. Financing talent rather than wealth directs resources toward more productive entrepreneurs, leading to higher output, faster growth, and greater upward mobility. At the same time, because returns become increasingly concentrated among highly productive entrepreneurs, talent-based financing also generates higher income and wealth inequality. The results therefore point to a fundamental trade-off between mobility, efficiency, and inequality that arises from the way entrepreneurial capital is allocated.

Overall, our findings suggest that measured genetic endowments are one source of persistent heterogeneity whose economic consequences depend critically on the environment in which they are expressed. Family investments, occupational choices, and financial institutions all shape how underlying differences translate into economic outcomes. As a result, policies that alter access to opportunities and the allocation of capital can affect not only aggregate efficiency, but also the degree to which individual talents are converted into entrepreneurship, mobility, and wealth accumulation across generations.

References

Abbott, B., Gallipoli, G., Meghir, C., and Violante, G. L. (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, 127(6):2569–2624.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R., and Hemous, D. (2019). Innovation and top income inequality. *Review of Economic Studies*.

Altonji, J. G. and Segal, L. M. (1996). Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics*, 14(3):353–366.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2017). Global inequality dynamics: New findings from wid. world. *American Economic Review*, 107(5):404–409.

Arold, B., Hufe, P., and Stoeckli, M. (2025). Genetic endowments, educational outcomes and the moderating influence of school investments. *Journal of Political Economy Microeconomics*. Advance online publication.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review*, 110(1):48–85.

Bach, L., Calvet, L. E., and Sodini, P. (2020). Rich pickings? risk, return, and skill in the portfolios of the wealthy. *American Economic Review*, 110(9):2703–2747.

Barban, N., Jansen, R., De Vlaming, R., Vaez, A., Mandemakers, J. J., Tropf, F. C., Shen, X., Wilson, J. F., Chasman, D. I., Nolte, I. M., et al. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature genetics*, 48(12):1462–1472.

Barth, D., Papageorge, N. W., and Thom, K. (2020). Genetic endowments and wealth inequality. *Journal of Political Economy*, 128(4):1474–1522.

Barth, D., Papageorge, N. W., Thom, K., and Velásquez-Giraldo, M. (2022a). Genetic endowments, income dynamics, and wealth accumulation over the lifecycle. Technical report, National Bureau of Economic Research.

Barth, D., Papageorge, N. W., Thom, K., and Velásquez-Giraldo, M. (2022b). Genetic endowments, income dynamics, and wealth accumulation over the lifecycle. Technical report, National Bureau of Economic Research.

Baselmans, B. M., Jansen, R., Ip, H. F., van Dongen, J., Abdellaoui, A., van de Weijer, M. P., Bao, Y., Smart, M., Kumari, M., Willemsen, G., et al. (2019). Multivariate genome-wide analyses of the well-being spectrum. *Nature genetics*, 51(3):445–451.

Becker, G. S. and Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy*, 87(6):1153–1189.

Becker, G. S. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics*, 4(3):S1–S39.

Becker, J., Burik, C. A., Goldman, G., Wang, N., Jayashankar, H., Bennett, M., Belsky, D. W., Karlsson Linnér, R., Ahlskog, R., Kleinman, A., et al. (2021). Resource profile and user guide of the polygenic index repository. *Nature human behaviour*, 5(12):1744–1758.

Benyamin, B., Pourcain, B., Davis, O. S., Davies, G., Hansell, N. K., Brion, M.-J., Kirkpatrick, R., Cents, R. A., Franić, S., Miller, M., et al. (2014). Childhood intelligence is heritable, highly polygenic and associated with fnbp1l. *Molecular psychiatry*, 19(2):253–258.

Berndt, S. I., Gustafsson, S., Mägi, R., Ganna, A., Wheeler, E., Feitosa, M. F., Justice, A. E., Monda, K. L., Croteau-Chonka, D. C., Day, F. R., et al. (2013). Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature genetics*, 45(5):501–512.

Biroli, P., Galama, T., Von Hinke, S., Van Kippersluis, H., Rietveld, C. A., and Thom, K. (2026). The economics and econometrics of gene–environment interplay. *Review of Economic Studies*, 93(1):144–180.

Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption inequality and partial insurance. *American Economic Review*, 98(5):1887–1921.

Bolt, U., French, E., Hentall-MacCuish, J., and O'Dea, C. (2024). Intergenerational altruism and transfers of time and money: a life cycle perspective. Technical report, IFS Working Papers.

Bolt, U., French, E., Warrier, V., Yang, Q., and Zhang, W. (2025). Genetic endowments and lifetime earnings: Understanding the mechanisms.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739.

Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., Boomsma, D. I., et al. (2020). Avoiding dynastic, assortative mating, and population stratification biases in mendelian randomization through within-family analyses. *Nature communications*, 11(1):3519.

Cagetti, M. and De Nardi, M. (2006). Entrepreneurship, frictions, and wealth. *Journal of political Economy*, 114(5):835–870.

Cagetti, M. and De Nardi, M. (2009). Estate taxation, entrepreneurship, and wealth. *American Economic Review*, 99(1):85–111.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The quarterly journal of economics*, 129(4):1553–1623.

Choi, S., Mak, T., and O'Reilly, P. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, 15(9):2759–2772.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Daruich, D. (2018). The macroeconomic consequences of early childhood development policies. *FRB St. Louis Working Paper*, (2018-29).

David, S., Hamidovic, A., Chen, G., Bergen, A., Wessel, J., Kasberger, J., Brown, W., Petruzella, S., Thacker, E., Kim, Y., et al. (2012). Genome-wide meta-analyses of smoking behaviors in african americans. *Translational psychiatry*, 2(5):e119–e119.

Day, F. R., Ong, K. K., and Perry, J. R. (2018). Elucidating the genetic basis of social interaction and isolation. *Nature communications*, 9(1):2457.

De Moor, M. H., Costa, P. T., Terracciano, A., Krueger, R. F., De Geus, E. J., Toshiko, T., Penninx, B. W., Esko, T., Madden, P. A., Derringer, J., et al. (2012). Meta-analysis of genome-wide association studies for personality. *Molecular psychiatry*, 17(3):337–349.

De Nardi, M. (2004). Wealth inequality and intergenerational links. *Review of Economic Studies*, 71(3):743–768.

Dearden, L., Machin, S., and Reed, H. (1997). Intergenerational mobility in britain. *The Economic Journal*, 107(440):47–66.

Deary, V., Hagenaars, S. P., Harris, S. E., Hill, W. D., Davies, G., Liewald, D. C., McIntosh, A. M., Gale, C. R., and Deary, I. J. (2018). Genetic contributions to self-reported tiredness. *Molecular psychiatry*, 23(3):609–620.

Del Boca, D., Flinn, C., and Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies*, 81(1):137–185.

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics*, 51(1):63–75.

DiPrete, T. A., Burik, C. A., and Koellinger, P. D. (2018). Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data. *Proceedings of the National Academy of Sciences*, 115(22):E4970–E4979.

Dunn, T. and Holtz-Eakin, D. (2000). Financial capital, human capital, and the transition to self-employment: Evidence from intergenerational links. *Journal of Labor Economics*, 18(2):282–305.

Durlauf, S. N. and Rustichini, A. (2023). Theory matters for identifying a causal role for genetic factors in socioeconomic outcomes. *Behavioral & Brain Sciences*, 46.

Evans, D. S. and Jovanovic, B. (1989). An estimated model of entrepreneurial choice under liquidity constraints. *Journal of political economy*, 97(4):808–827.

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

Fairlie, R. W. and Robb, A. M. (2007). *Why are black-owned businesses less successful than white-owned businesses? The role of families, inheritances, and business human capital*, volume 25. The University of Chicago Press.

Gabaix, X. and Landier, A. (2008). Why has ceo pay increased so much? *The quarterly journal of economics*, 123(1):49–100.

Gabaix, X., Lasry, J.-M., Lions, P.-L., and Moll, B. (2016). The dynamics of inequality. *Econometrica*, 84(6):2071–2111.

Gallipoli, G., Low, H., and Mitra, A. (2020). *Consumption and income inequality across generations*. Centre for Economic Policy Research.

Gourinchas, P.-O. and Parker, J. A. (2002). Consumption over the life cycle. *Econometrica*, 70(1):47–89.

Gregg, P., Macmillan, L., and Vittori, C. (2017). Moving towards estimating sons' lifetime intergenerational economic mobility in the uk. *Oxford bulletin of economics and statistics*, 79(1):79–100.

Guo, N. and Leung, C. K. Y. (2021). Do elite colleges matter? the impact on entrepreneurship decisions and career dynamics. *Quantitative Economics*, 12(4):1347–1397.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 20(1):25–46.

Hamilton, B. H. (2000). Does entrepreneurship pay? an empirical analysis of the returns to self-employment. *Journal of Political Economy*, 108(3):604–631.

Hamilton, B. H., Papageorge, N. W., and Pande, N. (2019). The right stuff? personality and entrepreneurship. *Quantitative Economics*, 10(2):643–691.

Harris, S. E., Hagenaars, S. P., Davies, G., David Hill, W., Liewald, D. C., Ritchie, S. J., Marioni, R. E., METASTROKE Consortium, I. C. f. B. P. G.-W. A. S., for Blood Pressure Genome-Wide Association Studies, I. C., Aging, C. C., Group, L., et al. (2017). Molecular genetic contributions to self-rated health. *International journal of epidemiology*, 46(3):994–1009.

Hassler, J., Rodríguez Mora, J. V., and Zeira, J. (2007). Inequality and mobility. *Journal of Economic Growth*, 12(3):235–259.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.

Hill, W. D., Hagenaars, S. P., Marioni, R. E., Harris, S. E., Liewald, D. C., Davies, G., Okbay, A., McIntosh, A. M., Gale, C. R., and Deary, I. J. (2016). Molecular genetic contributions to social deprivation and household income in uk biobank. *Current Biology*, 26(22):3083–3089.

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A., Gale, C. R., Davies, G., and Deary, I. J. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular psychiatry*, 24(2):169–181.

Holtz-Eakin, D., Joulfaian, D., and Rosen, H. S. (1994a). Entrepreneurial decisions and liquidity constraints. *RAND Journal of Economics*, 25(2):334–347.

Holtz-Eakin, D., Joulfaian, D., and Rosen, H. S. (1994b). Sticking it out: Entrepreneurial survival and liquidity constraints. *Journal of Political Economy*, 102(1):53–75.

Horikoshi, M., Beaumont, R. N., Day, F. R., Warrington, N. M., Kooijman, M. N., Fernandez-Tajes, J., Feenstra, B., Van Zuydam, N. R., Gaulton, K. J., Grarup, N., et al. (2016). Genome-wide associations for birth weight and correlations with adult disease. *Nature*, 538(7624):248–252.

Houmark, M. A., Ronda, V., and Rosholm, M. (2024). The nurture of nature and the nature of nurture: How genes and investments interact in the formation of skills. *American Economic Review*, 114(2):385–425.

Hout, M. and Rosen, H. S. (1999). Self-employment, family background, and race.

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., Coleman, J. R., Hagenaars, S. P., Ward, J., Wigmore, E. M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience*, 22(3):343–352.

Huggett, M., Ventura, G., and Yaron, A. (2011). Sources of lifetime inequality. *Journal of Political Economy*, 119(4):739–788.

Hurst, E. and Pugsley, B. W. (2011). What do small businesses do? *Brookings Papers on Economic Activity*, 42(2 (Fall)):73–142.

Hvide, H. K. and Oyer, P. (2018). Dinner table human capital and entrepreneurship. Technical report, National Bureau of Economic Research.

Jansen, P. R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A. R., de Leeuw, C. A., Benjamins, J. S., Muñoz-Manchado, A. B., Nagel, M., et al. (2019). Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature genetics*, 51(3):394–403.

Jolivet, G. and Postel-Vinay, F. (2025). A structural analysis of mental health and labour market trajectories. *Review of Economic Studies*, 92(3):1920–1954.

Jones, C. I. and Kim, J. (2018). A schumpeterian model of top income inequality. *Journal of Political Economy*, 126(5):1785–1826.

Jones, S. E., Tyrrell, J., Wood, A. R., Beaumont, R. N., Ruth, K. S., Tuke, M. A., Yaghootkar, H., Hu, Y., Teder-Laving, M., Hayward, C., et al. (2016). Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. *PLoS genetics*, 12(8):e1006125.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. M. (2019). The rate of return on everything, 1870–2015. *The quarterly journal of economics*, 134(3):1225–1298.

Kaplan, S. N. and Rauh, J. D. (2013). Family, education, and sources of wealth among the richest americans, 1982–2012. *American Economic Review*, 103(3):158–162.

Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., Lebreton, M., Tino, S. P., Abdellaoui, A., Hammerschlag, A. R., et al. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature genetics*, 51(2):245–257.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of political Economy*, 105(3):473–522.

Kihlstrom, R. E. and Laffont, J.-J. (1979). A general equilibrium entrepreneurial theory of firm formation based on risk aversion. *Journal of political economy*, 87(4):719–748.

Kiyotaki, N. and Moore, J. (1997). Credit cycles. *Journal of political economy*, 105(2):211–248.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., et al. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359(6374):424–428.

Koudstaal, M., Sloof, R., and Van Praag, M. (2016). Risk, uncertainty, and entrepreneurship: Evidence from a lab-in-the-field experiment. *Management Science*, 62(10):2897–2915.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*, 50(8):1112–1121.

Lee, S. Y. and Seshadri, A. (2019). On the intergenerational transmission of economic status. *Journal of Political Economy*, 127(2):000–000.

Levine, R. and Rubinstein, Y. (2017a). Smart and illicit: who becomes an entrepreneur and do they earn more? *The Quarterly journal of economics*, 132(2):963–1018.

Levine, R. and Rubinstein, Y. (2017b). Smart and illicit: Who becomes an entrepreneur and does it pay? *Quarterly Journal of Economics*, 132(2):963–1018.

Levine, R. and Rubinstein, Y. (2018). Selection into entrepreneurship and self-employment. Technical report, National Bureau of Economic Research.

Lindquist, M. J., Sol, J., and Van Praag, M. (2015). Why do entrepreneurial parents have entrepreneurial children? *Journal of Labor Economics*, 33(2):269–296.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., et al. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature genetics*, 51(2):237–244.

Luttmer, E. G. J. (2016). On the mechanics of firm growth. *Review of Economic Studies*, 83(1):58–109.

Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the united states using social security earnings data. *The Review of Economics and Statistics*, 87(2):235–255.

Mills, M. C., Tropf, F. C., Brazel, D. M., van Zuydam, N., Vaez, A., t Hoen Peter AC 62 63 van Meurs Joyce 66 Isaacs Aaron 72 Jansen Rick 34 Franke Lude 20 24, B. C. M. T. H. B. T. . ., van Meurs Joyce 66 Jhamai P. Mila 66 Verbiest Michael 66 Suchiman H. Eka D. 32 Verkerk Marijn 66 van der Breggen Ruud 32 van Rooij Jeroen 66 Lakenberg Nico 32, D. G., et al. (2021). Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nature human behaviour*, 5(12):1717–1730.

Nicolaou, N. and Shane, S. (2010). Entrepreneurship and occupational choice: Genetic and environmental influences. *Journal of Economic Behavior & Organization*, 76(1):3–14.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S. F. W., et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, 533(7604):539–542.

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., et al. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature genetics*, 54(4):437–449.

Pasman, J. A., Verweij, K. J., Gerring, Z., Stringer, S., Sanchez-Roige, S., Treur, J. L., Abdellaoui, A., Nivard, M. G., Baselmans, B. M., Ong, J.-S., et al. (2018). Gwas of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal effect of schizophrenia liability. *Nature neuroscience*, 21(9):1161–1170.

Plomin, R. and von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, 19(3):148–159.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38(8):904–909.

Quadrini, V. (1999). The importance of entrepreneurship for wealth concentration and mobility. *Review of income and Wealth*, 45(1):1–19.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., et al. (2013). Gwas of 126,559 individuals identifies genetic variants associated with educational attainment. *science*, 340(6139):1467–1471.

Rosen, S. (1981). The economics of superstars. *The American economic review*, 71(5):845–858.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Rustichini, A., Iacono, W. G., Lee, J. J., and McGue, M. (2023). Educational attainment and intergenerational mobility: a polygenic score analysis. *Journal of Political Economy*, 131(10):2724–2779.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics*, 50(7):912–919.

Schumann, G., Liu, C., O'Reilly, P., Gao, H., Song, P., Xu, B., Ruggeri, B., Amin, N., Jia, T., Preis, S., et al. (2016). Klb is associated with alcohol drinking, and its gene product β-klotho

is necessary for fgf21 regulation of alcohol preference. *Proceedings of the National Academy of Sciences*, 113(50):14372–14377.

Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., Preuss, M., Stewart, A. F., Barbalic, M., Gieger, C., et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*, 43(4):333–338.

Soler Artigas, M., Wain, L. V., Miller, S., Kheirallah, A. K., Huffman, J. E., Ntalla, I., Shrine, N., Obeidat, M., Trochet, H., McArdle, W. L., et al. (2015). Sixteen new lung function signals identified through 1000 genomes project reference panel imputation. *Nature communications*, 6.

Solon, G. (1992). Intergenerational income mobility in the united states. *American Economic Review*, 82(3):393–408.

Solon, G. (2004). A model of intergenerational mobility variation over time and place. *Generational Income Mobility in North America and Europe*, pages 38–47.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4):267–276.

Timmers, P. R., Mounier, N., Lall, K., Fischer, K., Ning, Z., Feng, X., Bretherick, A. D., Clark, D. W., Shen, X., et al. (2019). Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *elife*, 8:e39856.

Todd, P. E. and Zhang, W. (2020). A dynamic model of personality, schooling, and occupational choice. *Quantitative economics : journal of the Econometric Society.*, 11(1):231–275.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and Yang, J. (2017). 10 years of gwas discovery: biology, function, and translation. *American Journal of Human Genetics*, 101(1):5–22.

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., Aliev, F., Bacanu, S.-A., Batzler, A., Bertelsen, S., et al. (2018). Transancestral gwas of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature neuroscience*, 21(12):1656–1669.

Watson, H. J., Yilmaz, Z., Thornton, L. M., Hübel, C., Coleman, J. R., Gaspar, H. A., Bryois, J., Hinney, A., Leppä, V. M., Mattheisen, M., et al. (2019). Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nature genetics*, 51(8):1207–1214.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., Visscher, P. M., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of european ancestry. *Human molecular genetics*, 27(20):3641–3649.

Young, A. I., Nehzati, S. M., Benonisdottir, S., Okbay, A., Jayashankar, H., Lee, C., Cesarini, D., Benjamin, D. J., Turley, P., and Kong, A. (2022). Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nature genetics*, 54(6):897–905.

Appendices

A Genetic Endowments

This study measures genetic endowments using polygenic scores (PGSs) calculated from estimated effect sizes of genetic variants obtained through genome-wide association studies (GWAS).

The human genome comprises 23 pairs of chromosomes inherited from biological parents. These chromosomes, composed of DNA, carry genetic instructions encoded by nucleotide sequences: adenine (A), thymine (T), guanine (G), and cytosine (C). While most nucleotide sites are identical across humans, certain sites, called single-nucleotide polymorphisms (SNPs), vary across populations. Each SNP has a common allele and a less common allele. Individuals inherit two alleles at each SNP, one from each parent, resulting in SNP values of 0, 1, or 2 that indicate the number of less common alleles present.

To identify how SNPs relate to specific outcomes such as educational attainment, researchers conduct GWAS that regress outcomes on individual SNPs across large training samples, estimating an effect size (coefficient) for each SNP. The estimated coefficients, denoted by $\hat{\beta}_j$, are used to construct polygenic scores for each individual by summing weighted SNP values:

$$\widehat{PGS_i^y} = \sum_{j=1}^{J} \widehat{\beta_j^y} \cdot SNP_{ij}, \tag{A.1}$$

where PGS_i^y is the polygenic score for outcome y for individual i, and SNP_{ij} is the number of the less common alleles at SNP j.

Recent evidence (e.g., Okbay et al. (2016)) suggests that including all SNPs, rather than only significant ones, improves prediction accuracy. Following this approach, we construct PGSs using all SNPs to capture an individual's genetic endowments toward specific outcomes.

The Understanding Society dataset provides comprehensive genetic information from blood and saliva samples collected from 9,961 participants. The dataset includes 24,214,238 SNPs based on the hg38 genome build. Using PLINK2 and following the standard pipeline in Choi et al. (2020), We utilize publicly available GWAS summary statistics to construct polygenic scores for 48 traits among 9,920 individuals who pass quality control. These traits include educational attainment, risk tolerance, mental health, socioeconomic status (SES), physical composition, height, and weight.

Table A.1: GWAS Studies Used for Polygenic Score (PGS) Construction

Outcome	GWAS Study
Cognitive and Educational Outcomes	
Educational Attainment	Okbay et al. (2022)
Cognitive Performance	Savage et al. (2018)
Intelligence	Hill et al. (2019)
Childhood IQ	Benyamin et al. (2014)
Mental Health and Personality Traits	
ADHD	Demontis et al. (2019)
Mental Health	Baselmans et al. (2019)
Major Depressive Disorder (MDD)	Howard et al. (2019)
Tiredness	Deary et al. (2018)
Loneliness	Day et al. (2018)
Agreeableness	De Moor et al. (2012)
Anorexia Nervosa	Watson et al. (2019)
Risk-related Behaviors	
Age at First Sex	Mills et al. (2021)
Speeding Propensity	Karlsson Linnér et al. (2019)
Smoking / Ever Smoked	David et al. (2012); Karlsson Linnér et al. (2019)
Risk Tolerance / Risk Behavior Composite	Karlsson Linnér et al. (2019)
Drinks Per Week / Alcohol Use	Karlsson Linnér et al. (2019); Schumann et al. (2016)
Cannabis Use	Pasman et al. (2018)
Cigarettes Per Day	Liu et al. (2019)
Alcohol Dependence	Walters et al. (2018)
Physical Health and Well-being	
Self-Rated Health	Harris et al. (2017)
Coronary Artery Disease	Schunkert et al. (2011)
Obesity / BMI	Berndt et al. (2013); Yengo et al. (2018)
Forced Expiratory Volume (FEV)	Soler Artigas et al. (2015)
Insomnia / Sleep Duration / Chronotype	Jansen et al. (2019); Jones et al. (2016)
Socioeconomic Status (SES) and Demographic Outcomes	
Household Income	Hill et al. (2016)
Age at First Birth	Barban et al. (2016)
Longevity	Timmers et al. (2019)
Number of Children Ever Born	Barban et al. (2016)
Townsend Deprivation Index	Hill et al. (2016)
Height/Weight	
Birth Weight	Horikoshi et al. (2016)
Height	Yengo et al. (2018)

Notes: This table summarizes the source Genome-Wide Association Studies (GWAS) used to construct the Polygenic Scores (PGS) employed in our analysis. The selection of studies prioritizes large-scale meta-analyses with European-ancestry samples to ensure maximal predictive power. Each PGS is calculated as a weighted sum of risk alleles, where weights correspond to the effect sizes estimated in the respective discovery GWAS listed above. For outcomes with multiple references, weights are derived from the most recent or comprehensive meta-analysis available.

A.1 Measurement Error in Genetic Measure and Its Correction

Polygenic scores are constructed from estimated SNP coefficients and therefore inherit sampling noise from the underlying GWAS. This section combines the split-sample measurement-error logic in (Becker et al., 2021) and (DiPrete et al., 2018) with the errors-in-variables argument used in (Heckman et al., 2013) to derive a plug-in measurement-error correction for genetic regressors and the corresponding HC1 heteroskedasticity-robust covariance matrix. Throughout, variables are assumed to be demeaned, so sample cross-products coincide with centered second moments.

Setup. Let the latent genetic measure be

$$G_i = \sum_{s=1}^{S} b_s a_{is}, \tag{A.2}$$

where a_{is} denotes the allele count at SNP s and b_s is the true SNP effect. The observed score uses estimated SNP effects $\hat{b}_s$ and can therefore be written as

$$\tilde{G}_i = \sum_{s=1}^{S} \hat{b}_s a_{is} = G_i + \nu_i, \tag{A.3}$$

where ν_i is the measurement error induced by sampling noise in the GWAS coefficients. Under the classical measurement-error restriction,

$$\mathbb{E}[\nu_i \mid \Omega_{G,i}, Z_i, \varepsilon_i] = 0, \qquad \nu_i \perp\!\!\!\perp (\Omega_{G,i}, Z_i, \varepsilon_i), \tag{A.4}$$

the outcome equation of interest is

$$y_i = \beta_g \Omega_{G,i} + Z_i' \gamma + \varepsilon_i. \tag{A.5}$$

where $\Omega_{G,i}$ is set of genetic measures of interest and Z_i' is the control.

Split-sample correction for the latent genetic measures. Following (DiPrete et al., 2018) and (Bolt et al., 2025), we construct two genetic measures from independent GWAS training samples, A and B, where A denotes the primary sample.[24] Let

$$\tilde{G}_{A,i} = G_i + \nu_{A,i}, \qquad \tilde{G}_{B,i} = \lambda G_i + \nu_{B,i}, \tag{A.6}$$

where the measurement errors are mutually orthogonal and orthogonal to the true measure.

[24]The primary sample is the GWAS training sample that shares the same population background as our analytical sample. For an analysis conducted in the UK setting, the GWAS training sample from the UK Biobank is the primary sample.

To recover the loading factor λ, let H_i denote any phenotype orthogonal to the measurement errors. Then

$$\lambda = \frac{\text{Cov}(\tilde{G}_{B,i}, H_i)}{\text{Cov}(\tilde{G}_{A,i}, H_i)}, \tag{A.7}$$

which implies we can estimate the variance of the latent genetic measure as follows:

$$\text{Var}(G_i) = \frac{\text{Cov}(\tilde{G}_{A,i}, \tilde{G}_{B,i})}{\lambda}. \tag{A.8}$$

Let $\overline{G^p}$ be the mean parental genetic measure

$$\overline{G_i^p} = \frac{1}{2}\left(G_i^m + G_i^f\right),$$

By Mendelian segregation, the Mendelian Draw is orthogonal to the parental genetic measure:

$$(G_i - \overline{G_i^p}) \perp\!\!\!\perp \overline{G_i^p}, \tag{A.9}$$

Thus variance-covariance matrix of $\Omega_{G,i} = (G_i - \overline{G_i^p}, \overline{G_i^p})'$ is:

$$\begin{aligned}\Sigma_{\Omega_G\Omega_G} &= \begin{pmatrix} \Sigma_{(G-\overline{G^p})(G-\overline{G^p})} & \Sigma_{(G-\overline{G^p})\overline{G^p}} \\ \Sigma_{\overline{G^p}(G-\overline{G^p})} & \Sigma_{\overline{G^p}\overline{G^p}} \end{pmatrix} = \begin{pmatrix} \Sigma_{(G-\overline{G^p})(G-\overline{G^p})} & 0 \\ 0 & \Sigma_{\overline{G^p}\overline{G^p}} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{(G-\overline{G^p})(G-\overline{G^p})} & 0 \\ 0 & \Sigma_{GG} - \Sigma_{(G-\overline{G^p})(G-\overline{G^p})} \end{pmatrix}\end{aligned} \tag{A.10}$$

where $\Sigma_{(G-\overline{G^p})(G-\overline{G^p})}$ and Σ_{GG} can be obtained based on the method illustrated in equations (A.6) to (A.8).

Errors-in-variables estimator. Let $\tilde{X}_i = (\tilde{\Omega}_{G,i}, Z_i')'$ and $X_i = (\Omega_{G,i}, Z_i')'$, and collect the regression coefficients in $\beta = (\beta_g, \gamma')'$. Under classical measurement error assumption (A.4) the probability limit of the OLS estimator is:

$$\begin{aligned}\text{plim}\,\hat{\beta}_{\text{OLS}} &= \begin{pmatrix} \Sigma_{\tilde{\Omega}_G\tilde{\Omega}_G} & \Sigma_{\tilde{\Omega}_G Z} \\ \Sigma_{Z\tilde{\Omega}_G} & \Sigma_{ZZ} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{\Omega_G\Omega_G} & \Sigma_{\Omega_G Z} \\ \Sigma_{Z\Omega_G} & \Sigma_{ZZ} \end{pmatrix} \beta \\ &= \underbrace{\begin{pmatrix} \Sigma_{\tilde{\Omega}_G\tilde{\Omega}_G} & \Sigma_{\tilde{\Omega}_G Z} \\ \Sigma_{Z\tilde{\Omega}_G} & \Sigma_{ZZ} \end{pmatrix}^{-1}}_{M} \underbrace{\begin{pmatrix} \Sigma_{\Omega_G\Omega_G} & \Sigma_{\tilde{\Omega}_G Z} \\ \Sigma_{Z\tilde{\Omega}_G} & \Sigma_{ZZ} \end{pmatrix}}_{\tilde{X}'\tilde{X}} \beta.\end{aligned} \tag{A.11}$$

which indicates the OLS estimates are inconsistent. Follow (Heckman et al., 2013), we construct the EIV (error-in-variable) consistent estimates as follow:

$$\hat{\beta}_{\text{EIV}} = M^{-1}(\tilde{X}'\tilde{X})^{-1}\hat{\beta}_{OLS} = M^{-1}\tilde{X}'y. \tag{A.12}$$

Since $M/n \xrightarrow{p} \mathbb{E}[X_i X_i']$ and $\tilde{X}'y/n \xrightarrow{p} \mathbb{E}[X_i y_i]$, thus

$$\operatorname{plim} \hat{\beta}_{\text{EIV}} = \left(\mathbb{E}[X_i X_i']\right)^{-1} \mathbb{E}[X_i y_i] = \beta. \tag{A.13}$$

HC1 covariance matrix for the EIV estimator. Let $\hat{u}_i$ be fitted residuals using the EIV estimates:

$$\hat{u}_i = y_i - \tilde{X}_i' \hat{\beta}_{\text{EIV}}, \tag{A.14}$$

and let k denote the total number of coefficients. The HC1 covariance matrix is:

$$\widehat{\Omega} = \frac{n}{n-k} \operatorname{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2), \tag{A.15}$$

$$\widehat{\operatorname{Var}}_{\text{HC1}}(\hat{\beta}_{\text{EIV}}) = \widehat{M}^{-1} \tilde{X}' \widehat{\Omega} \tilde{X} \widehat{M}^{-1\prime}. \tag{A.16}$$

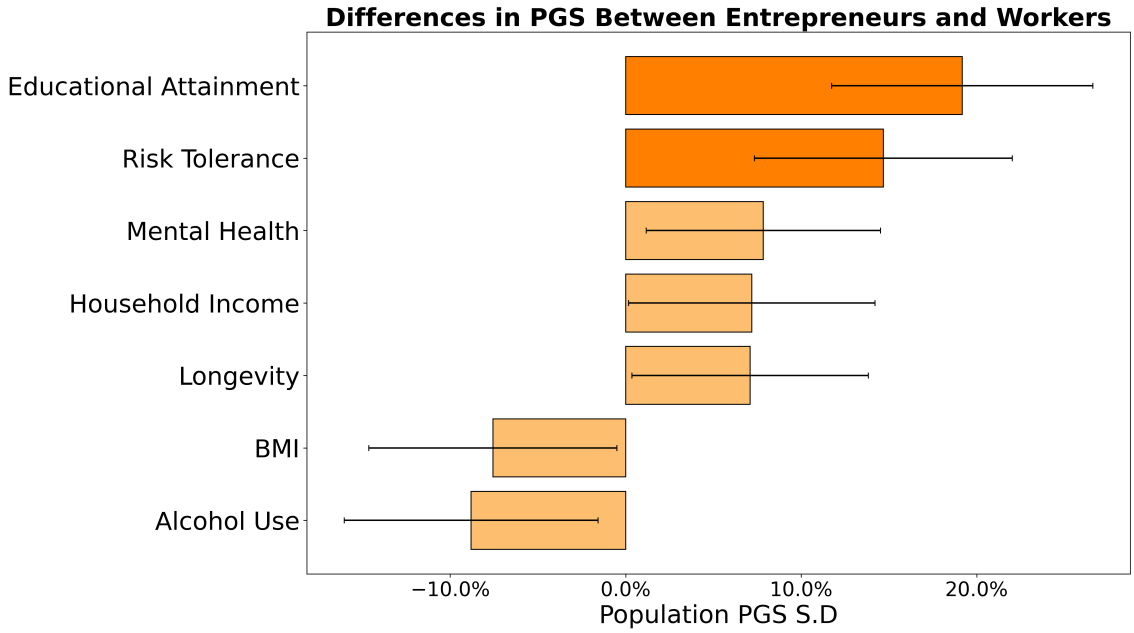
Accordingly, the HC1 standard error for coefficient j is

$$\widehat{\text{SE}}_{\text{HC1},j} = \sqrt{\left[\widehat{\operatorname{Var}}_{\text{HC1}}(\hat{\beta}_{\text{EIV}})\right]_{jj}}. \tag{A.17}$$

A.2 Distribution of Polygenic Scores

This appendix documents two features of the polygenic score distributions used in the analysis. First, it shows how mean PGS values differ between individuals ever observed as entrepreneurs and those always observed in paid employment. Second, it examines the empirical relationship between PGS_{edu} and PGS_{risk}, the two genetic measures used in the structural model.

Figure A.1: Differences in PGS Distributions Between Entrepreneurs and Workers

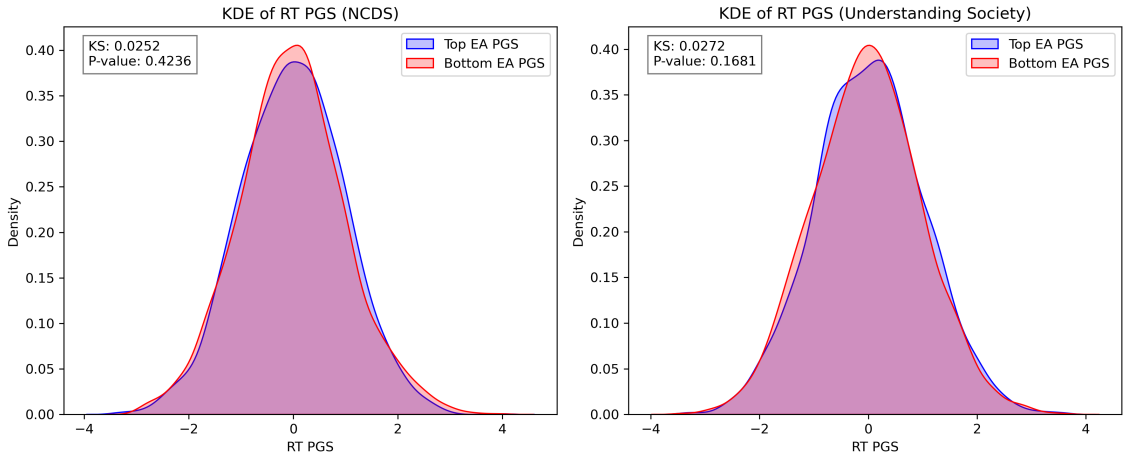


Notes: This figure compares mean polygenic scores between individuals who are ever observed as entrepreneurs and individuals always observed in paid employment. Bars report differences in group means, with 95% confidence intervals. The traits include educational attainment, risk tolerance, income, personality, mental health, and longevity.

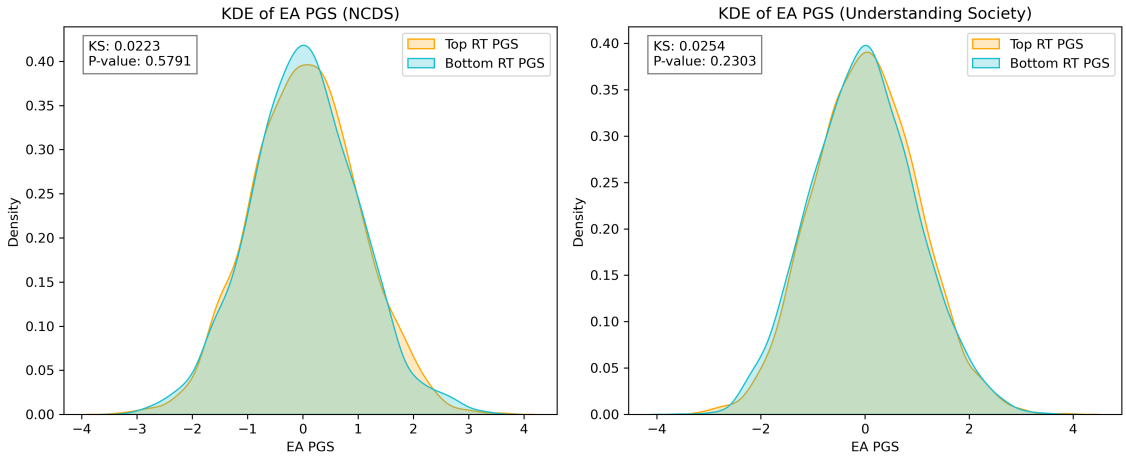
Figure A.1 reports raw mean differences in polygenic scores between entrepreneurs and workers. This descriptive comparison complements the controlled logistic estimates in Figure 1 in the main text. Consistent with the main empirical results, educational attainment and risk tolerance exhibit the largest positive differentials for entrepreneurs.

Because the model uses PGS_{edu} and PGS_{risk} as two distinct dimensions of genetic endowment, we next examine whether the two scores capture empirically separate variation. Figure A.2 compares the distribution of each PGS across high and low values of the other PGS. The distributions are visually similar, and the Kolmogorov–Smirnov tests do not reject equality of distributions. This supports treating educational-attainment and risk-tolerance PGSs as separate dimensions of genetic heterogeneity in the model.

Figure A.2: Orthogonality Tests of Polygenic Scores Across Datasets



(a) PGS_{risk} Distribution by Top and Bottom PGS_{edu} Groups



(b) PGS_{edu} Distribution by Top and Bottom PGS_{risk} Groups

Notes: This figure examines the empirical relationship between PGS_{edu} and PGS_{risk} in the NCDS and Understanding Society datasets. Panel (a) plots the kernel density of PGS_{risk} separately for individuals in the top and bottom groups of the PGS_{edu} distribution. Panel (b) plots the kernel density of PGS_{edu} separately for individuals in the top and bottom groups of the PGS_{risk} distribution. Kolmogorov–Smirnov test statistics and p-values are reported in each panel.

B Data

This study utilizes data from Understanding Society (the UK Household Longitudinal Study), which builds upon the British Household Panel Survey (BHPS) initiated in 1991 and integrated into Understanding Society in 2009. This nationally representative longitudinal panel tracks individuals annually, providing comprehensive information on household income, employment status, housing conditions, family structure, health, education, risk attitudes, and well-being. Particularly relevant for this research is the genetic subsample, comprising 9,961 individuals who provided blood samples for DNA extraction. This genetic data enables the construction of individual-specific polygenic scores (PGSs), which serve as measures of genetic endowments and can be linked to diverse economic and behavioral outcomes. Table B.1 summarizes the genetic sample in Understanding Society. Two key advantages of this dataset support the investigation of genetic effects on entrepreneurial choice within a dynastic framework.

First, the richness of Understanding Society data facilitates comprehensive analysis of labor-market and intertemporal-choice variables. Among the 9,919 individuals in the genetic sample who pass quality control, 9,647 report self-employment status, 9,915 report educational attainment, and 7,178 report saving behavior. Crucially, a core analytical subsample of 7,029 individuals possesses complete data on education, self-employment, and savings, enabling rigorous analysis of these outcomes in relation to genetic endowments.

Second, the dataset's capacity to observe genetic relationships within families facilitates analysis of intergenerational transmission of genetic endowments from parents to offspring. A total of 1,299 child–parent pairs have available genetic data: 550 individuals have genetic information on a biological parent, 392 on a biological child, and 223 on a biological sibling. Additionally, the dataset contains genetic endowment information for spouses or partners: 2,213 genotyped spouses and 3,509 genotyped partners. These features enable investigation of mechanisms such as genetic nurture and assortative mating, providing a more comprehensive understanding of how genetics influence economic outcomes across generations.

This combination of individual-level genetic data with rich economic and demographic information across generations and family members makes Understanding Society a unique resource for studying both the direct and indirect effects of genetic endowments on economic behavior.

Table B.1: Sample Size by Information Availability

Sample Category	**N**
Core Analysis Sample	
Full genetic sample (post-QC)	9,919
Self-employment information	9,647
Educational attainment	9,915
Saving information	7,178
Child educational attainment	5,930
Linked Biological Relatives	
Biological parent genotype linked	550
Biological child genotype linked	392
Biological sibling genotype linked	223

Notes: This table summarizes the sample sizes available for key variables after applying standard genetic Quality Control (QC) protocols. *Full genetic sample* refers to individuals with valid genomic data. *Linked* categories indicate successful matching of genotype records between the primary respondent and their respective biological relatives within the dataset.

B.1 Measures of Skills and Parental Investments

Table B.2: NCDS Measures of Skills, Parental Investment

Panel A: Skill measures

Timing	Cognitive skills	Non-cognitive skills
Age 7	Reading score; maths score; draw-a-man score; copying score	Inconsequential; hostile to children; writes off adult standards; hostile to adults; miserable/depressed
Age 11	Reading score; maths score; copying score	Inconsequential; hostile to children; writes off adult standards; hostile to adults; miserable/depressed
Age 16	Reading score; maths score; teacher-rated English ability; teacher-rated maths ability	Unconcentrated; fights frequently; disobedient; irritable; miserable/depressed

Panel B: Parental investment measures

Timing	Time investment measures	School quality measures
Age 7	Teacher's assessment of interest in education (mother and father); outings with child (mother and father); read to child (mother and father)	Whether schools conduct parent-teacher meetings; social class of fathers of children in class; whether school has a parent-teacher association; type of school; number of children in class who have had a parent-teacher meeting
Age 11	Teacher's assessment of parents' interest in education (mother and father); outings with child (mother and father); father's involvement in upbringing; parents' ambitions regarding child's educational attainment (further education and university)	Student-teacher ratio; type of school; number of children in class suitable for GCEs
Age 16	Teacher's assessment of parents' interest in education; involvement of parents in child's schooling	Number of children in class studying for GCEs; type of school; percentage of children staying past school-leaving age; student-teacher ratio

Notes: Data are collected at ages 7, 11, and 16 in the NCDS. Investment measures are retrospective: investments from ages 0–7 are measured at age 7, investments from ages 7–11 are measured at age 11, and investments from ages 11–16 are measured at age 16.

B.2 Employment Classification

We classify each person-wave observation using three types of information from the Understanding Society genetic sample: (i) whether the respondent is self-employed or works for an employer; (ii)

for self-employed respondents, whether the individual reports running a business or professional practice (or being a partner in such a practice), or operating informally as a freelancer, sub-contractor, sole trader, or other self-employed; and (iii) for employees, whether the position is full-time (30 or more hours per week) or part-time. Table B.3 lists the resulting categories and the inclusion decision applied throughout the analysis.

We retain two groups. Entrepreneurs are self-employed individuals who report running a business or professional practice, or being a partner in such a practice (3,825 observations). Workers include full-time employees (47,850 observations) and unincorporated self-employed individuals (freelancers, sub-contractors, sole traders, and other self-employed), whose activity reflects flexible labor supply rather than business ownership (6,490 observations). We drop three categories that cannot be cleanly assigned: self-employed respondents who do not specify the nature of their work (220 observations); employees who do not report full-time or part-time status (1,617 observations); and part-time employees (15,963 observations), whose constrained labor supply is inconsistent with primary career commitment.

Table B.3: Employment Classification

Category	Description	N	Decision
Self-employed, type unclassified	Respondent reports self-employment but does not specify the nature of work	220	Dropped
Entrepreneur	Respondent reports running a business or professional practice, or being a partner in such a practice	3,825	Kept as Entrepreneur
Unincorporated self-employed	Respondent reports working for themselves as a freelancer, sub-contractor, sole trader, or other self-employed	6,490	Kept as Worker
Employee, hours unclassified	Respondent reports working for an employer but does not specify full-time or part-time	1,617	Dropped
Full-time employee	Respondent reports working for an employer for 30 or more hours per week	47,850	Kept as Worker
Part-time employee	Respondent reports working for an employer for fewer than 30 hours per week	15,963	Dropped

Notes: N denotes person-wave observations across all survey waves. Entrepreneurs and workers together form the estimation sample used throughout the analysis.

B.3 Parental Genotype Imputation

All 1,299 individuals in the reduced-form sample have parental polygenic scores imputed using the SNIPAR procedure of Young et al. (2022). The key idea is to use genetic relatedness within families to recover the component of genotype inherited from the parents, even when one or both parents are not directly genotyped. What differs across individuals is not whether imputation is used, but how much family information is available to identify the parental component, and therefore how precise the resulting mid-parental PGS measure is.

The sample contains three family configurations. First, 578 individuals have at least one genotyped parent in the Understanding Society/BHPS sample. In these cases, the observed

parental genotype provides strong information for the imputation, so the resulting mid-parental PGS is measured relatively precisely. Second, 648 individuals are themselves parents with at least one genotyped child in the sample, but without genotyped parents. For these individuals, SNIPAR uses identity-by-descent (IBD) sharing between the individual and the child to infer the grandparental genetic contribution, which in turn identifies the parental component relevant for the individual. Because the grandparental haplotype cannot be fully assigned without directly observing the grandparents, the procedure recovers a symmetric mid-parental estimate, which is less precise but remains unbiased. Third, 73 individuals are observed only in sibling pairs. In these cases, SNIPAR uses IBD sharing between siblings to recover information about the parental genotype.

C Robustness Check on Empirical Evidence

C.1 Genetic Effect on Entrepreneurship: Joint Polygenic-Score Specification

In the main text we estimate equation (2.1) one polygenic score at a time, recovering the marginal association of each genetic predisposition with entrepreneurial entry. Because polygenic scores are genetically correlated, a natural question is whether the two leading predictors, PGS_{edu} and PGS_{risk}, retain independent explanatory power once the scores are allowed to compete. To assess this, we re-estimate the model with all polygenic scores entered simultaneously:

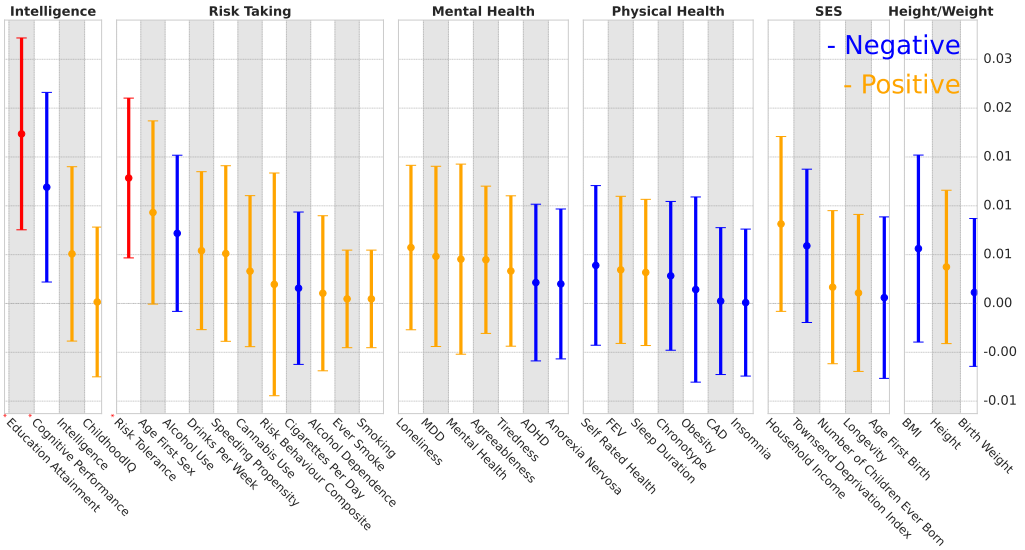
$$\Pr(\text{Entrepreneur}_i = 1) = \Lambda\left(\beta_0 + \sum_{k=1}^{K} \beta_k \,\text{PGS}_{ik} + \beta'_{PC}\,\text{PC}_i + \beta_S\,\text{SEX}_i + \epsilon_i\right), \tag{C.1}$$

where PGS_{ik} denotes the k-th of the K polygenic scores for individual i, and the controls are as in equation (2.1).

The interpretation of the coefficients changes accordingly. Whereas α_1 in the single-score specification captures the total association of a given score, each β_k now captures its partial association, that is, the predictive content of PGS_{ik} orthogonal to the remaining scores. Because genetically correlated scores share variance, and because each score is a noisy proxy for its latent trait, the joint estimates are mechanically attenuated relative to their marginal counterparts. They should be read as a conservative lower bound on each score's contribution rather than as a revised measure of its total association.

Figure C.1 reports the joint estimates. The central finding is unchanged. PGS_{edu} and PGS_{risk} retain their sign, remain statistically significant, and continue to rank as the two strongest predictors of entrepreneurial entry. That both survive mutual adjustment indicates the main-text results are not an artifact of genetic correlation among the scores. Each predisposition carries independent information about who becomes an entrepreneur, consistent with the two capturing distinct channels (human capital and risk tolerance) rather than a single shared genetic dimension.

Figure C.1: Joint Associations Between PGSs and the Probability of Ever Being an Entrepreneur



Notes: Standardized coefficients are estimated from a single logistic regression of ever being an entrepreneur on all polygenic scores (PGSs) entered jointly, per equation (C.1). The regression controls for gender and the top 10 genetic principal components. PGSs are grouped into six domains: intelligence, mental health, risk-related behavior, physical health, socioeconomic status (SES), and height/weight. Each bar displays the absolute value of the estimated association, with horizontal lines indicating 95% confidence intervals. Yellow bars denote positive associations and blue bars negative ones. The red bar highlights PGS_{edu}. Statistically significant estimates are marked with red asterisks. Abbreviations: MDD = major depressive disorder; CAD = coronary artery disease; FEV = forced expiratory volume (lung function).

C.2 Intergenerational Occupational Persistence: Biological vs. Adopted Children

To provide quasi-experimental evidence for the genetic channel, we exploit the distinction between biological and adopted children. If genes matter, persistence should be stronger among biological children, who share them with their parents, than among adopted children, who share only the family environment. We estimate:

$$y_i = \alpha_{f(i)} + \beta\, p_i + \delta\, A_i + \theta\,(p_i \times A_i) + X_i'\Omega + \varepsilon_i \tag{C.2}$$

where y_i equals one if child i was ever observed in the large employer class, p_i indicates whether at least one parent held the same class during the child's upbringing, A_i indicates adopted child status, X_i includes sex, birth-decade, and region fixed effects, and $\alpha_{f(i)}$ is either a common intercept (columns 1–3) or a family fixed effect (column 4). Columns (1) and (2) estimate β separately on the biological and adopted subsamples. Column (3) pools both groups and estimates equation (C.2) in full sample; θ measures the gap in persistence between biological and adopted children. Column (4) restricts to mixed families in which the same parent raised both biological and adopted children; since p_i is constant within a family, $\beta\, p_i$ and $\theta\,(p_i \times A_i)$ are absorbed into $\alpha_{f(i)}$, and δ is identified entirely from variation between biological and adopted children raised under the same roof. Equation (C.2) answers the question: do biological children reach the parent's occupational class at higher rates than their adopted siblings in the same

household? A positive estimate implicates genes; a zero points to the environment.

Table C.1: Parental Occupation and Child Occupational Attainment: Biological vs. Adopted Children

	(1) Bio only	(2) Adopted only	(3) Full sample	(4) Mixed families
Parent: large employer ($\hat{\beta}$)	0.0243**	−0.0411	0.0243**	
	(0.0117)	(0.0437)	(0.0116)	
Adopted ($\hat{\delta}$)			0.0068	0.0105
			(0.0111)	(0.0131)
Parent × adopted ($\hat{\theta}$)			−0.0408**	
			(0.0170)	
Family FE	No	No	No	Yes
N	8,342	185	8,527	229
R^2	0.018	0.263	0.026	0.228

Notes: LPM = linear probability model. Child outcome: whether the child was ever observed in the large employers & higher management occupational class. Parent occupation is measured from the childhood module, reflecting the respondent's retrospective report of their parent's occupation when growing up. All columns estimate equation (C.2); $\alpha_{f(i)}$ is a common intercept in columns (1)–(3) and a family fixed effect in column (4). Columns (1) and (2) estimate $\hat{\beta}$ separately for biological and adopted children. Column (3) pools both groups and adds the interaction term. Column (4) restricts to mixed families, defined as parents with both biological and adopted children; since p_i is constant within a family, $\hat{\beta}$ and $\hat{\theta}$ are absorbed into $\mu_{f(i)}$, and $\hat{\delta}$ is identified from within-family variation in adoption status, with standard errors clustered by family. All columns include sex, birth-decade, and region fixed effects except column (4). Full results across occupational classes are reported in Table C.2 below. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.1 reports the results. Among biological children, having a parent in the large employer class raises the probability of reaching the same class by 2.4 percentage points. Given that only about 1.4% of children ever reach this class, the implied persistence is economically large. The same coefficient for adopted children is −4.1 percentage points and statistically indistinguishable from zero, pointing to no detectable environmental transmission of large-employer status. The interaction term in column (3) confirms this gap is significant ($\hat{\theta} = -0.041$, $p < 0.05$): the persistence observed among biological children essentially disappears for adopted children raised in the same types of households. The within-family estimate $\hat{\delta}$ in column (4) is consistent with the cross-sectional pattern but cannot reject zero on its own [25]. Taken together, the evidence points to a meaningful genetic component in the transmission of entrepreneurship. A more comprehensive comparison across occupational classes is reported in Table C.2 below.

[25]The estimate is insignificant, as the mixed-family sample is small ($N = 229$).

Table C.2: Childhood Parent Occupation and Child Occupational Attainment: Biological vs. Adopted Children

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Large employers & higher management				Higher professional				Intermediate occupations			
	Bio only	Adopted only	Full sample	Mixed families	Bio only	Adopted only	Full sample	Mixed families	Bio only	Adopted only	Full sample	Mixed families
Parent: large employer ($\hat{\beta}$)	0.0243**	−0.0411	0.0243**									
	(0.0117)	(0.0437)	(0.0116)									
Parent: higher professional ($\hat{\beta}$)					0.0965***	−0.0422	0.0966***					
					(0.0197)	(0.0378)	(0.0197)					
Parent: intermediate occ. ($\hat{\beta}$)									0.0599***	−0.2014*	0.0598***	
									(0.0155)	(0.1133)	(0.0155)	
Adopted ($\hat{\delta}$)			0.0068	0.0105			−0.0173	0.0633*			0.0591*	0.0172
			(0.0111)	(0.0131)			(0.0132)	(0.0373)			(0.0327)	(0.0883)
Parent × Adopted ($\hat{\theta}$)			−0.0408**				−0.1397***				−0.1304	
			(0.0170)				(0.0258)				(0.1307)	
Family FE	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
N	8,342	185	8,527	229	8,342	185	8,527	229	8,342	185	8,527	229
R^2	0.018	0.263	0.026	0.228	0.026	0.133	0.026	0.163	0.048	0.259	0.050	0.109

Notes: LPM (Linear Probability Model). Child outcome for each column group: whether the child was ever observed in the respective occupational class: large employers & higher management, higher professional, or intermediate occupations. Parent occupation is measured from the childhood module, reflecting the respondent's retrospective report of their parent's occupation when growing up. All columns estimate equation (C.2); $\alpha_{f(i)}$ is a common intercept in columns (1)–(3), (5)–(7), and (9)–(11), and a family fixed effect in columns (4), (8), and (12). In the mixed-family columns, since p_i is constant within a family, $\hat{\beta}$ and $\hat{\theta}$ are absorbed into $\mu_{f(i)}$, and $\hat{\delta}$ is identified from within-family variation in adoption status. Columns (1)–(3), (5)–(7), and (9)–(11) include sex, birth-decade, and region fixed effects with HC1 robust standard errors. Columns (4), (8), and (12) use standard errors clustered by family. Stars: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.3 Comparison of OLS and EIV Estimates

Table C.3: Reduced-Form Evidence: Entrepreneurial Intention and Entrepreneurial Entry

	EIV			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Ever expected to start a business (N = 1,276)						
Child's PGS_{edu} deviation	0.011	0.021	0.020	0.006	0.011	0.010
	(0.011)	(0.013)	(0.013)	(0.006)	(0.007)	(0.007)
Child's PGS_{risk} deviation	0.018*	0.024**	0.024**	0.012*	0.017**	0.017**
	(0.010)	(0.012)	(0.012)	(0.007)	(0.008)	(0.008)
Mid-parental PGS_{edu}	–	−0.006	−0.011	–	−0.000	−0.002
		(0.024)	(0.025)		(0.009)	(0.008)
Mid-parental PGS_{risk}	–	0.006	0.003	–	0.006	0.004
		(0.021)	(0.021)		(0.010)	(0.010)
Family SES controls	–	–	✓	–	–	✓
Panel B: Entrepreneurial entry (N = 1,299)						
Child's PGS_{edu} deviation	0.043***	0.032*	0.028	0.022***	0.019*	0.017*
	(0.016)	(0.019)	(0.019)	(0.008)	(0.010)	(0.010)
Child's PGS_{risk} deviation	0.006	0.004	0.004	0.004	0.003	0.003
	(0.011)	(0.014)	(0.014)	(0.008)	(0.010)	(0.010)
Mid-parental PGS_{edu}	–	0.066**	0.059*	–	0.026**	0.023**
		(0.033)	(0.033)		(0.011)	(0.011)
Mid-parental PGS_{risk}	–	0.011	0.007	–	0.006	0.004
		(0.024)	(0.024)		(0.011)	(0.011)
Family SES controls	–	–	✓	–	–	✓

Notes: This table reports reduced-form estimates using a linear probability model. Panel A reports estimates for entrepreneurial intention, where the dependent variable is a binary indicator for whether the individual ever responds "Yes" to "Do you expect to start your own business?" in any survey wave. Panel B reports estimates for entrepreneurial entry, where the dependent variable is a binary indicator for whether the individual ever becomes an entrepreneur. Default controls include sex, the top ten genetic principal components, survey-wave fixed effects, and missing dummies. Family SES controls comprise parental occupational-class indicators; the parental education index is excluded. The left block reports EIV (error-in-variables) estimates adjusting for measurement error in the genetic variables, and the right block reports OLS estimates. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table C.4: Reduced-Form Evidence: Log Earnings

	EIV			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Entrepreneur log earnings (N = 432)						
Child's PGS_{edu} deviation	−0.018	−0.003	0.147	−0.010	−0.016	0.001
	(0.172)	(0.213)	(0.400)	(0.064)	(0.084)	(0.083)
Child's PGS_{risk} deviation	0.261**	0.344***	0.249	0.167**	0.236***	0.234***
	(0.115)	(0.119)	(0.416)	(0.073)	(0.075)	(0.075)
Mid-parental PGS_{edu}	–	0.246	1.792	–	0.020	0.073
		(0.972)	(5.143)		(0.104)	(0.122)
Mid-parental PGS_{risk}	–	−0.351	−0.588	–	−0.052	−0.037
		(0.411)	(0.929)		(0.118)	(0.122)
Family SES controls	–	–	✓	–	–	✓
Panel B: Worker log earnings (N = 8,290)						
Child's PGS_{edu} deviation	0.165***	0.111***	0.109***	0.083***	0.071***	0.068***
	(0.013)	(0.015)	(0.015)	(0.006)	(0.008)	(0.007)
Child's PGS_{risk} deviation	0.008	0.022*	0.021*	0.005	0.012	0.011
	(0.009)	(0.012)	(0.012)	(0.007)	(0.008)	(0.008)
Mid-parental PGS_{edu}	–	0.259***	0.236***	–	0.101***	0.092***
		(0.029)	(0.028)		(0.009)	(0.009)
Mid-parental PGS_{risk}	–	−0.027	−0.033	–	−0.007	−0.012
		(0.021)	(0.021)		(0.010)	(0.010)
Family SES controls	–	–	✓	–	–	✓

Notes: This table reports reduced-form estimates for log annual gross earnings. Panel A reports entrepreneur log annual earnings. Panel B reports worker log annual earnings. Default controls include sex, the top ten genetic principal components, age, and age squared. Family SES controls comprise parental occupational-class indicators; the parental education index is excluded. The left block reports EIV (error-in-variables) estimates adjusting for measurement error in the genetic variables, and the right block reports OLS estimates. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

C.4 Sample Selection and Entropy Balancing

Background. The 1,299-individual related-imputation sample is selected because at least one parent, child, or sibling is present in the Understanding Society/BHPS panel, enabling mid-parental PGS imputation. This criterion creates demographic imbalances relative to the full 9,919-person analytic sample. The most consequential imbalance is age: the related-imputation sample is substantially younger, reflecting the fact that older cohorts are less likely to have a living, genotyped parent in the panel. The sample also has somewhat more years of education and slightly higher self-reported risk preference, consistent with its younger age composition.

Entropy balancing procedure. To address these imbalances, we apply entropy balancing (Hainmueller, 2012), which reweights observations in the related-imputation sample to exactly satisfy moment constraints with respect to the full-sample distribution. Formally, the method solves the dual problem

$$\min_{\gamma} \; \log\left(\sum_{i=1}^{N_s} e^{\mathbf{x}_i'\gamma}\right) \; - \; \bar{\mathbf{x}}_{\text{full}}'\gamma,$$

where $\mathbf{x}_i$ is the vector of balancing variables for related-imputation-sample observation i and $\bar{\mathbf{x}}_{\text{full}}$ is the full-sample mean vector. The optimal weights are $w_i^* = N_s \cdot \text{softmax}(\mathbf{x}_i'\hat{\gamma})$, with mean equal to 1.

Choice of balancing variables. We balance on 15 variables: age, an indicator for female, the first ten genetic principal components, years of education (with a missing indicator), and self-reported risk preference (with a missing indicator). Age and age squared address the main demographic imbalance. Education and risk preference are included because they are correlated with both the probability of having a genotyped relative and with the outcomes of interest; omitting them could leave residual imbalance even after balancing on age. The genetic principal components guard against population stratification in the reweighting.

Balance achieved. Table C.5 reports the SMD (standardised mean difference) for key variables before and after entropy balancing. Before balancing, the related-imputation sample is substantially younger than the full sample. After balancing, all SMDs are below 10^{-6}, confirming exact satisfaction of the moment constraints. The effective sample size after reweighting is approximately 731.

Table C.5: Balance Statistics Before and After Entropy Balancing

	Standardized mean difference	
Variable	Before balancing	After balancing
Age	−0.67	$< 10^{-6}$
Female	0.00	$< 10^{-6}$
Years of education	0.11	$< 10^{-6}$
Risk preference	0.16	$< 10^{-6}$
Genetic PC 1	−0.12	$< 10^{-6}$
Genetic PC 2	0.06	$< 10^{-6}$

Notes: SMD = (related-imputation-sample mean − full-sample mean) / full-sample SD. The balancing set includes age, age squared, female, genetic PCs 1–10, years of education, an education missing indicator, risk preference, and a risk preference missing indicator (16 variables total). All 16 SMDs satisfy $|\text{SMD}| < 10^{-6}$ after balancing. Effective sample size after balancing: $N_{\text{eff}} \approx 731$ out of $N = 1{,}299$.

Robustness. Table C.6 compares the individual PGS estimates from the main specification (equation (2.2), column 1) in the unweighted related-imputation sample and in the entropy-balanced related-imputation sample. The direction of each estimate is preserved under entropy balancing across all outcomes. For entry, the PGS_{edu} signal becomes smaller and loses conventional significance, consistent with the smaller effective sample size rather than a reversal of the underlying relationship. For entrepreneur earnings, a one-standard-deviation higher PGS_{risk} is associated with a somewhat larger earnings premium after reweighting, roughly one-third higher earnings. This confirms that the risk-tolerance earnings gradient is not an artifact

of the younger age composition of the related-imputation sample. The qualitative conclusions from the main specification carry over to the entropy-balanced sample in every outcome.

Table C.6: Individual PGS Estimates: Unweighted and Entropy-Balanced Samples

	Unweighted		Entropy-balanced	
Outcome	PGS_{edu}	PGS_{risk}	PGS_{edu}	PGS_{risk}
Ever expects to start a business	0.009	0.017**	0.001	0.013
	(0.006)	(0.008)	(0.006)	(0.008)
Ever run a business	0.006	0.009*	0.003	0.007*
	(0.004)	(0.005)	(0.003)	(0.004)
Entrepreneur log earnings	−0.174*	0.351***	−0.202*	0.363***
	(0.096)	(0.078)	(0.108)	(0.087)
Wage worker log earnings	0.064***	0.013	0.076***	0.008
	(0.007)	(0.008)	(0.009)	(0.009)
Sample size	1,299		Eff. $N \approx 731$	

Notes: Each row reports individual PGS coefficients from equation (2.2), column (1) specification. Controls include sex, genetic PCs 1–10, and wave fixed effects; age and age squared are added for earnings. Standard errors (HC1) are reported in parentheses for unweighted estimates. Entropy-balanced estimates use WLS with weights from Appendix C.4. Significance: $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$. Earnings are restricted to ages 24–60.

C.5 Other Regression Results

Table C.7: Genetic Effects on Education and Risk-Taking Phenotypes

	Age left school	Risk taking
Child PGS_{edu}	0.263***	−0.109
	(0.060)	(0.132)
Child PGS_{risk}	0.014	0.428***
	(0.059)	(0.116)
Father PGS_{edu}	0.130*	−0.139
	(0.079)	(0.178)
Mother PGS_{edu}	0.082	0.100
	(0.073)	(0.163)
Father PGS_{risk}	0.047	0.035
	(0.069)	(0.154)
Mother PGS_{risk}	−0.014	0.172
	(0.069)	(0.149)
N	1,291	549

Notes: Dependent variables are age left school and self-reported risk taking. Regressors of interest are the standardized child, father, and mother PGS components for education and risk tolerance, entered jointly. All columns include sex, age, top 10 genetic principal components, and wave fixed effects. HC1 standard errors are reported in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table C.8: Outcomes on Education and Risk Taking

Panel A: Entrepreneurial intention and entry

	Entrepreneurial intention	Successful entry
Age left school	0.012***	0.003***
	(0.001)	(0.001)
Risk taking	0.006***	0.002***
	(0.001)	(0.001)
R^2	0.021	0.009
N	9,591	9,591

Panel B: Log earnings

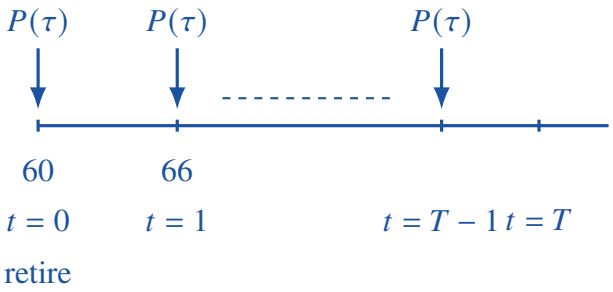
	Worker log earnings	Entrepreneurial log earnings
Age left school	0.123***	0.143***
	(0.002)	(0.014)
Risk taking	0.015***	0.034**
	(0.002)	(0.015)
R^2	0.169	0.078
N	47,483	2,847

Notes: Regressors of interest are years of education and risk taking, with missing indicators included in every column. Panel A columns include sex and wave fixed effects. Panel B columns include sex, age, and age squared. HC1 standard errors are reported in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

D Other Model and Identification Details

D.1 Terminal Value for Retired Agents

Timing. The retired agent enters retirement with assets a_{retire} and type τ. Pension income is type-specific, $P(\tau)$, and the first pension payment arrives at time zero. We set the terminal age to 102 ($T = 7$):



Problem. For a fixed pre-market type τ, let the period utility be $u(c, \tau) = \frac{c^{1-\gamma_\tau}}{1-\gamma_\tau}$. The value function is given by

$$V_t(a_t, \tau) = \max_{c_t, a_{t+1}} \{u(c_t, \tau) + \beta V_{t+1}(a_{t+1}, \tau)\}$$

subject to

$$a_{t+1} = (1 + r)a_t + P(\tau) - c_t, \qquad a_T = 0.$$

Closed-Form Solution. The Euler equation yields

$$c_{t+1} = [\beta(1+r)]^{1/\gamma_\tau}\, c_t.$$

The present-value budget constraint is

$$a_{\text{retire}} + P(\tau) \sum_{i=0}^{T-1} \left(\frac{1}{1+r}\right)^i = \sum_{i=0}^{T-1} \left(\frac{1}{1+r}\right)^i c_i.$$

Therefore, we can solve for the consumption sequence analytically:

$$c_0(a_{\text{retire}}, \tau) = \frac{a_{\text{retire}} + P(\tau) \sum_{i=0}^{T-1} (1+r)^{-i}}{\sum_{i=0}^{T-1} \left(\frac{[\beta(1+r)]^{1/\gamma_\tau}}{1+r}\right)^i},$$

$$c_i(a_{\text{retire}}, \tau) = [\beta(1+r)]^{i/\gamma_\tau}\, c_0(a_{\text{retire}}, \tau), \qquad i = 0, \ldots, T-1.$$

Terminal value at age 60. The retirement value at $t = 0$ is given by

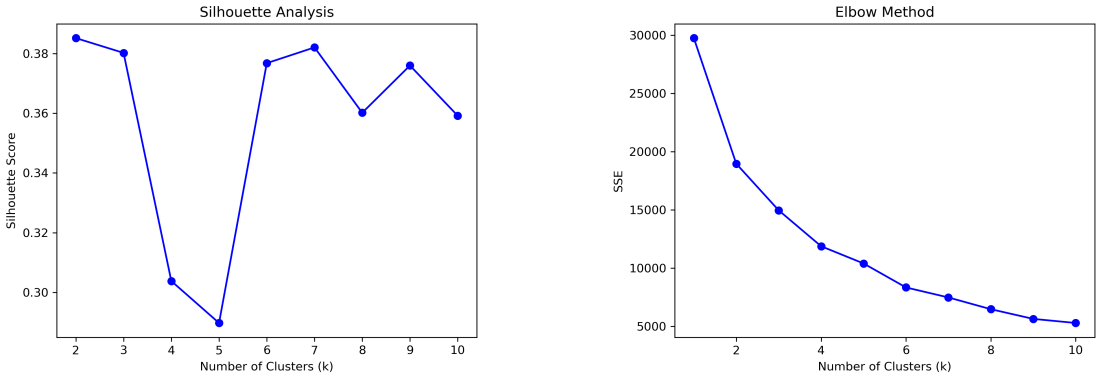
$$\phi(a_{\text{retire}}, \tau) = \frac{c_0(a_{\text{retire}}, \tau)^{1-\gamma_\tau}}{1-\gamma_\tau} \sum_{i=0}^{T-1} \beta^i\, [\beta(1+r)]^{i(1-\gamma_\tau)/\gamma_\tau}.$$

D.2 Details on Pre-market Type Identification

To identify distinct pre-market types within the sample, we employ a clustering approach based on pre-market characteristics. The optimal number of clusters is determined using two standard diagnostic criteria: the Silhouette Analysis (Rousseeuw, 1987) and the Elbow Method (Thorndike, 1953).

As illustrated in Figure D.1, both methods yield consistent results. Panel (a) shows the average silhouette width, a measure of how similar an object is to its own cluster compared to other clusters. The score peaks sharply at $k = 2$, indicating that a two-type configuration maximizes within-cluster cohesion and between-cluster separation. This is further supported by the Elbow Method in Panel (b), which plots the within-cluster sum of squares (WCSS) against the number of clusters. A distinct "elbow" point is observed at two clusters, after which the marginal reduction in WCSS becomes negligible. Consequently, we categorize the population into two latent groups: “High” and “Low” pre-market types.

Figure D.1: Optimal Cluster Number Determination



(a) Silhouette Analysis

(b) Elbow Method

Notes: This figure presents the diagnostic plots for determining the optimal number of clusters. Panel (a) displays the average silhouette width, where a higher value indicates better cluster separation. Panel (b) shows the within-cluster sum of squares (WCSS) using the elbow method. Both metrics identify $k = 2$ as the optimal cluster count.

D.3 Model Fit

Table D.1: Model Fit: Targeted Moments and Simulation Results

Moment	Simulated	Target
Panel A: Earnings and Financial Moments		
Entrepreneur earnings		
HH	12.72	11.66
HL	10.31	10.56
LH	11.04	10.02
LL	8.60	7.93
High type	12.80	12.75
Low type	7.61	7.37
Mean	10.96	10.16
Std. dev.	11.00	10.34
Autocorr.	0.56	0.51
Worker earnings		
HH	7.54	7.76
HL	5.74	7.44
LH	6.24	6.63
LL	4.85	6.45
High type	7.46	8.70
Low type	4.75	5.32
Mean	6.02	7.09
Std. dev.	5.18	4.27
Autocorr.	0.84	0.83
Financial moments		
Transfer	1.190	1.300
Saving ratio	0.091	0.129
Saving age 60	0.142	0.157
Consumption volatility ratio	0.936	0.920
Income Gini	0.343	0.330
Panel B: Entrepreneurship and Transitions		
Entrepreneurship share		
HH	0.135	0.124
HL	0.111	0.106
LH	0.122	0.114
LL	0.088	0.091
High type	0.139	0.138
Low type	0.090	0.084
Worker parent → entrepreneur child	0.109	0.124
Entrepreneur parent → entrepreneur child	0.196	0.190
Panel C: Investment and Type Formation		
Parental investment		
HH	0.172	0.205
HL	0.228	0.201
LH	−0.221	−0.170
LL	−0.206	−0.263
Pr(child high type), parent high type		
Parent HH	0.667	0.631
Parent HL	0.624	0.676
Parent LH	0.574	0.551
Parent LL	0.491	0.556
Pr(child high type), parent low type		
Parent HH	0.492	0.500
Parent HL	0.468	0.475
Parent LH	0.343	0.338
Parent LL	0.273	0.331

Notes: Earnings are measured in 2015 £10,000s. HH, HL, LH, and LL denote high/low combinations of PGS_{edu} and PGS_{risk}. The consumption volatility ratio is $[\text{Var}(\log c \mid \tau = H)/\text{Var}(\log y \mid \tau = H)]/[\text{Var}(\log c \mid \tau = L)/\text{Var}(\log y \mid \tau = L)]$.

Table D.2: Estimated Structural Parameters

Parameter	Estimate	Std. Err.
Entrepreneurial Ability Process		
$\beta_{0,E}$: Baseline log entrepreneurial ability	-0.4069	(0.0251)
$\beta_{g,E}^{HH}$: HH genetic premium	0.3811	(0.0261)
$\beta_{g,E}^{HL}$: HL genetic premium	0.1685	(0.0274)
$\beta_{g,E}^{LH}$: LH genetic premium	0.3424	(0.0258)
$\beta_{\tau,E}^{H}$: High pre-market type premium	0.4310	(0.0218)
Worker Wage Process		
$\beta_{0,W}$: Baseline log wage	0.8054	(0.0179)
$\beta_{g,W}^{HH}$: HH genetic premium	0.3669	(0.0199)
$\beta_{g,W}^{HL}$: HL genetic premium	0.1092	(0.0200)
$\beta_{g,W}^{LH}$: LH genetic premium	0.2962	(0.0233)
$\beta_{\tau,W}^{H}$: High pre-market type premium	0.5383	(0.0058)
Idiosyncratic Ability and Wage Shocks		
ρ_E: Entrepreneurial ability persistence	0.3550	(0.0125)
σ_E: Entrepreneurial ability shock std. dev.	0.6072	(0.0088)
ρ_W: Wage shock persistence	0.9192	(0.0052)
σ_W: Wage shock std. dev.	0.4523	(0.0027)
Intergenerational Transmission		
β: Discount factor	0.9079	(0.0081)
ρ_E^c: Intergenerational persistence of entrepreneurial ability	0.8592	(0.0312)
ρ_W^c: Intergenerational persistence of wage shock	0.0182	(0.0009)
Production Technology and Preferences		
ν: Return to entrepreneurial capital	0.6244	(0.0154)
λ: Altruism weight	0.0788	(0.0012)
γ_H: Risk aversion, high type	2.9718	(0.0154)
γ_L: Risk aversion, low type	3.4464	(0.0283)
ξ: Marginal utility cost of parental investment	3.64×10^{-3}	(1.39×10^{-4})
b_H: Non-pecuniary value of entrepreneurship, high type	9.58×10^{-4}	(9.9×10^{-5})
b_L: Non-pecuniary value of entrepreneurship, low type	-4.62×10^{-3}	(5.87×10^{-4})
Child Type Formation Logit		
$\delta_{g^c}^{HH}$: Child HH genetic group	0.3897	(0.0390)
$\delta_{g^c}^{HL}$: Child HL genetic group	0.3787	(0.0530)
$\delta_{g^c}^{LH}$: Child LH genetic group	0.0860	(0.0315)
δ_{τ}^{H}: Parent high type	1.2576	(0.0354)
δ_{g}^{HH}: Parent HH genetic group	0.6885	(0.0416)
δ_{g}^{HL}: Parent HL genetic group	0.5447	(0.0274)
δ_{g}^{LH}: Parent LH genetic group	0.3562	(0.0562)
δ_{ι^p}: Parental investment	1.4729	(0.0293)
$\delta_{HH}^{\iota^p}$: Investment × child HH	10.9141	(1.0092)
$\delta_{HL}^{\iota^p}$: Investment × child HL	1.3366	(0.1088)
$\delta_{LH}^{\iota^p}$: Investment × child LH	2.6541	(1.2146)
δ_0: Baseline log-odds	-1.6044	(0.0358)
Macro Aggregates		
α: Corporate capital share	0.2220	–
A: Corporate TFP	3.5639	–

Notes: Standard errors are reported in parentheses and are computed using a finite-difference numerical Jacobian. Estimates are from the SMM/GMM estimation. Earnings-process parameters are normalized relative to the low pre-market type and the LL genetic group. Type-formation parameters are normalized relative to a child in the LL genetic group with a low-type parent in the parental LL genetic group. Genetic groups are defined as: HH = high PGS_{edu}/high PGS_{risk}; HL = high PGS_{edu}/low PGS_{risk}; LH = low PGS_{edu}/high PGS_{risk}; LL = low PGS_{edu}/low PGS_{risk}.

D.4 Estimated Structural Parameters

D.5 Consumption-Equivalent Value of Non-Pecuniary Entrepreneurship Returns

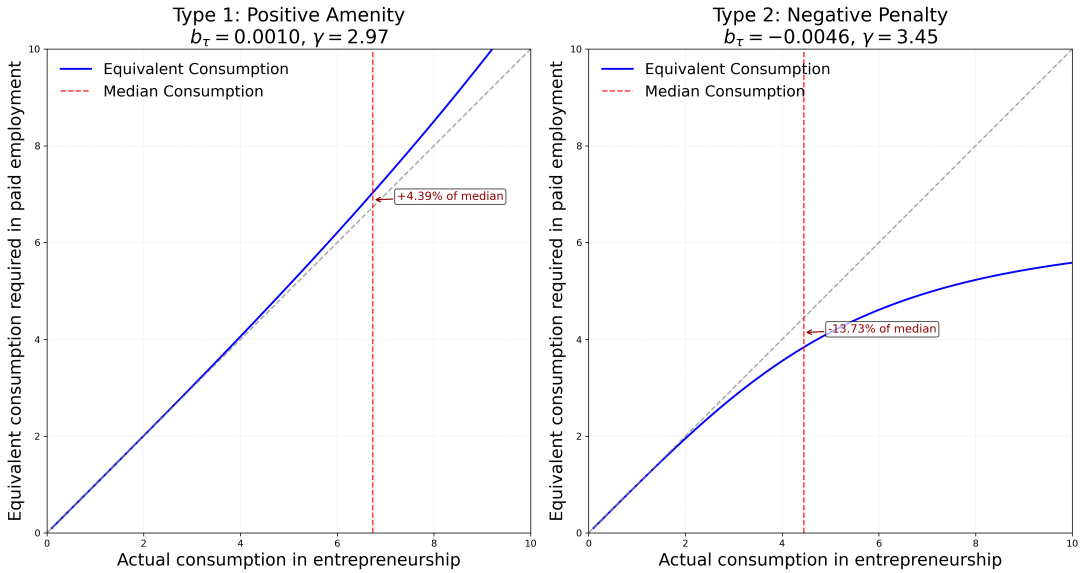
We translate the estimated non-pecuniary value of entrepreneurship into consumption-equivalent units. For each pre-market type τ, we compute the level of paid-employment consumption that gives the same flow utility as entrepreneurship consumption c_E:

$$\frac{c_W(c_E,\tau)^{1-\gamma_\tau}}{1-\gamma_\tau} = \frac{c_E^{1-\gamma_\tau}}{1-\gamma_\tau} + b_\tau.$$

The resulting $c_W(c_E,\tau)$ is the consumption required in paid employment to make the individual indifferent to entrepreneurship, holding pecuniary earnings fixed.

Figure D.2 plots this mapping for high and low pre-market types. The 45-degree line corresponds to the case with no non-pecuniary value. For high types, the equivalent-consumption curve lies above the 45-degree line, reflecting a positive amenity value from entrepreneurship. At median consumption, this amenity is equivalent to a 4.4 percent consumption gain. For low types, the curve lies below the 45-degree line, reflecting a non-pecuniary penalty from entrepreneurship. At median consumption, this penalty is equivalent to 13.7 percent of consumption. These calculations show that preference heterogeneity is economically meaningful, even though the estimated b_τ parameters are small in utility units.

Figure D.2: Consumption-Equivalent Value of Non-Pecuniary Entrepreneurship Returns



Notes: The figure plots the paid-employment consumption level that gives the same flow utility as entrepreneurship consumption, holding pecuniary consumption fixed. The dashed gray line is the 45-degree line. The red vertical line marks median consumption. Positive deviations from the 45-degree line indicate a non-pecuniary amenity from entrepreneurship; negative deviations indicate a non-pecuniary penalty.

E Additional Details for Counterfactual Analysis

E.1 Additional Wealth Distribution Decompositions

This appendix reports additional diagnostics for the wealth-distribution counterfactuals discussed in Section 6.4. The main text shows that genetic heterogeneity has limited effects on the main body of the wealth distribution but substantially affects the upper tail. Here, we first visualize the full wealth distribution under each counterfactual, and then provide two complementary decompositions. First, we report mean wealth ratios by percentile bin. Second, we decompose total wealth variance into within-top, within-bottom, and between-group components.

Figure E.1 visualizes how genetic mechanisms reshape the wealth distribution across counterfactuals. Above the top 1% of the wealth distribution, the benchmark economy exhibits a pronounced right tail. Removing the intergenerational genetic link compresses this tail, and removing individual genetic heterogeneity compresses it further. By contrast, the distribution for the bottom 99% changes only modestly across counterfactuals.

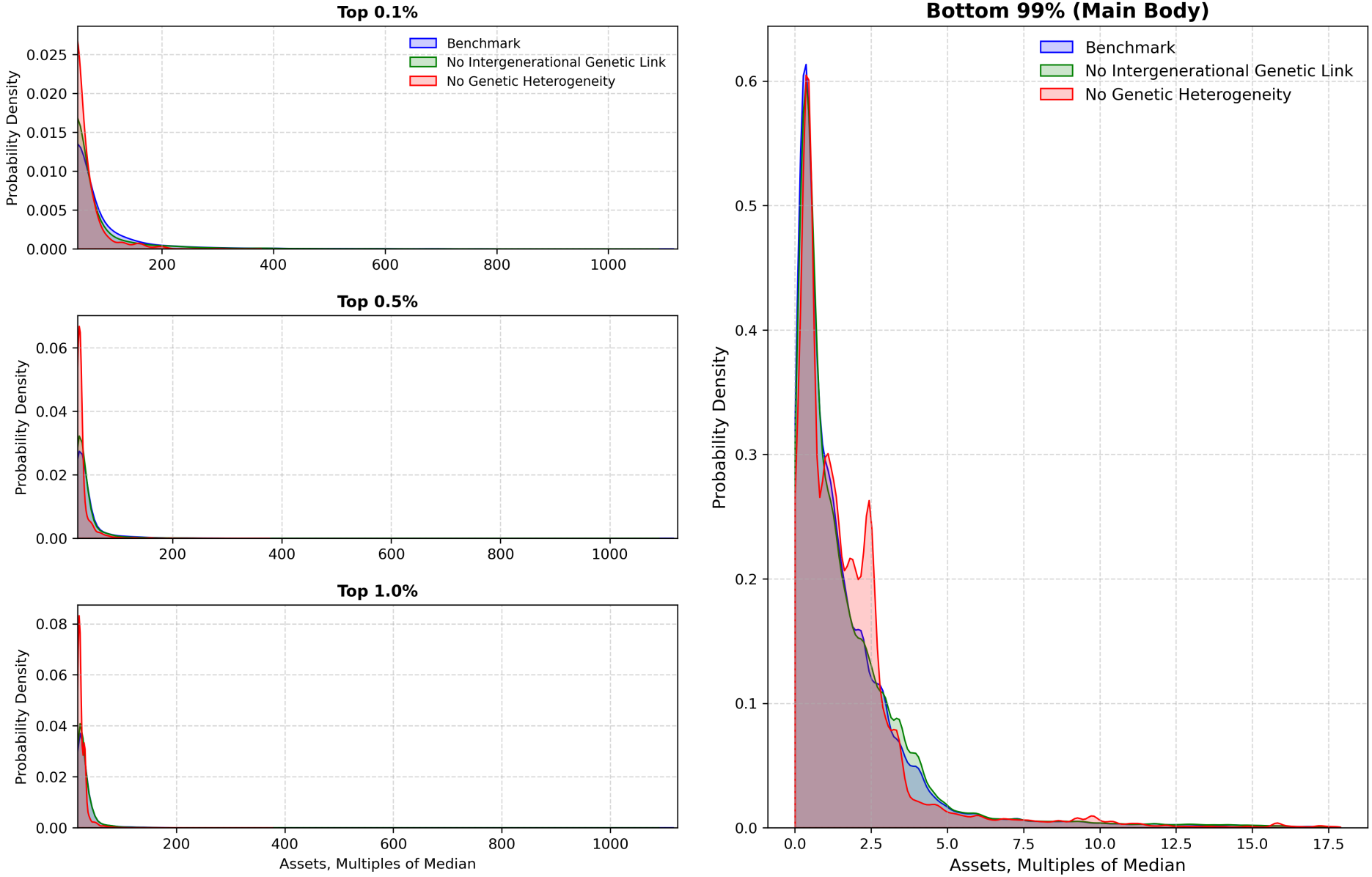
Table E.1 quantifies the change in mean wealth ratios by percentile bin. In the benchmark economy, households in the top 0.1% hold average wealth equal to 118 times the mean wealth of the bottom 90%. Removing the intergenerational genetic link reduces this ratio to 89, and removing genetic heterogeneity entirely reduces it to 45. The effect becomes progressively smaller as we move down the wealth distribution. For the top 1–10% bin, the total decline from Model 1 to Model 3 is less than 10%. Thus, the impact of genetic heterogeneity is highly non-linear: it is modest for most of the distribution but large at the very top.

Table E.1: Wealth Ratios by Percentile Bin Across Model Specifications

Percentile bin	Model 1 Benchmark	Model 2 No IG Link	Δ_{12}	Model 3 No G Heterogeneity	Δ_{23}	Δ_{13}
Top 0.1%	118.33	88.62	−25.1%	45.27	−36.6%	−61.7%
Top 0.1–0.5%	29.16	28.36	−2.7%	23.03	−18.3%	−21.0%
Top 0.5–1%	18.81	18.58	−1.2%	16.92	−8.8%	−10.0%
Top 1–10%	6.18	6.06	−2.0%	5.59	−7.7%	−9.6%
Bottom 90%	1.00	1.00	0.0%	1.00	0.0%	0.0%

Notes: The table reports mean wealth ratios by percentile bin. Ratios are expressed relative to the mean wealth of the bottom 90% within each model, so the bottom-90% bin is normalized to one. Model 1 is the benchmark economy with full genetic heterogeneity and intergenerational genetic transmission. Model 2 removes the intergenerational genetic link, so children draw genetic endowments independently of parental genetic endowments. Model 3 removes individual genetic heterogeneity. The columns Δ_{12}, Δ_{23}, and Δ_{13} report changes in the ratio when moving from Model 1 to Model 2, from Model 2 to Model 3, and from Model 1 to Model 3, respectively, expressed as percentages of the Model 1 ratio.

Figure E.1: Wealth Distribution under Different Genetic Transmission Mechanisms



Notes: This figure presents kernel density estimates of the wealth distribution across four wealth quantile segments. The left panel shows three tail distributions conditional on assets exceeding the 99.9th (top 0.1%), 99.5th (top 0.5%), and 99.0th (top 1%) percentiles. The right panel shows the distribution for the bottom 99% of the population. The horizontal axis reports assets normalized by the population median in the benchmark economy. The blue density corresponds to the benchmark model. The green density represents the counterfactual with no intergenerational genetic link. The red density represents the counterfactual with no genetic heterogeneity.

The mean ratios show that genetic mechanisms matter most at the top, but they do not show whether this effect comes from a larger gap between the top and the rest of the population or from greater dispersion among households within the top itself. We therefore separate these two margins in two steps.

First, within each model, we decompose total wealth variance into within-group and between-group components. We use the benchmark top 1% wealth cutoff and let T denote households above this cutoff and B the remaining 99%. For each model $m \in \{1, 2, 3\}$ defined in Figure 9, let $a_{i,m}$ denote wealth, μ_m the population mean, $p_{s,m}$ the population share of group $s \in \{T, B\}$, $\mu_{s,m}$ the group mean, and $\sigma^2_{s,m}$ the within-group variance. Total wealth variance can be written as

$$\mathrm{Var}_m(a) = \underbrace{p_{T,m}\sigma^2_{T,m}}_{C^T_m} + \underbrace{p_{B,m}\sigma^2_{B,m}}_{C^B_m} + \underbrace{p_{T,m}(\mu_{T,m} - \mu_m)^2 + p_{B,m}(\mu_{B,m} - \mu_m)^2}_{C^{TB}_m}.$$

Here C^T_m is the within-top 1% contribution, C^B_m is the within-bottom-99% contribution, and C^{TB}_m is the between-group contribution. Collect these components in the vector

$$C_m \equiv \begin{pmatrix} C^T_m \\ C^B_m \\ C^{TB}_m \end{pmatrix}, \qquad \mathrm{Var}_m(a) = \mathbf{1}' C_m.$$

Second, we decompose the role of genetic mechanisms by comparing this vector across the three counterfactual economies. We measure the total genetic contribution as the vector Δ^{AllG}, comparing each component of C_1 in the benchmark economy with the corresponding component of C_3 in the economy without genetic heterogeneity:

$$\Delta^{AllG} = \frac{C_1 - C_3}{\mathrm{Var}_1(a)}.$$

Similarly, we define vector Δ^{IG} as the contribution of the intergenerational genetic link and vector Δ^{OwnG} as the contribution of individual genetic heterogeneity:

$$\Delta^{IG} = \frac{C_1 - C_2}{\mathrm{Var}_1(a)}, \qquad \Delta^{OwnG} = \frac{C_2 - C_3}{\mathrm{Var}_1(a)}.$$

All components are normalized by benchmark total wealth variance, so the entries in Table E.2 can be read as percentage-point contributions to benchmark wealth dispersion.

Table E.2 shows that the genetic contribution to wealth dispersion is concentrated almost entirely within the top 1%. Genetic mechanisms account for 84.3 percentage points of benchmark wealth variance, of which 75.7 percentage points come from within-top 1% dispersion. The bottom 99% contributes only 0.5 percentage points. The intergenerational genetic link is the dominant component: it accounts for 58.0 percentage points of benchmark wealth variance through within-top 1% dispersion, compared with 17.8 percentage points from individual genetic heterogeneity. This shows that the wealth-distribution effect of genetic heterogeneity is concen-

Table E.2: Decomposition of Genetic Contributions to Wealth Variance

Component	IG link (Δ^{IG})	Own G (Δ^{OwnG})	Total G (Δ^{AllG})
Within bottom 99% (C^B)	−0.2%	0.7%	0.5%
Within top 1% (C^T)	58.0%	17.8%	75.7%
Between top 1% and bottom 99% (C^{TB})	1.8%	6.4%	8.1%
Total wealth variance	59.5%	24.9%	84.3%

Notes: Entries report contributions as shares of benchmark total wealth variance. The top 1% is defined using the benchmark wealth cutoff and held fixed across counterfactual economies. The IG-link contribution is $\Delta^{IG} = (C_1 - C_2)/\mathrm{Var}_1(a)$, the own-$G$ contribution is $\Delta^{OwnG} = (C_2 - C_3)/\mathrm{Var}_1(a)$, and the total genetic contribution is $\Delta^{AllG} = (C_1 - C_3)/\mathrm{Var}_1(a)$. Negative values indicate that removing the corresponding genetic channel increases that component relative to the benchmark.

trated at the top rather than spread evenly across the population, and that the intergenerational genetic link accounts for most of this effect.

E.2 Venture Capital Counterfactual: Mathematical Details

Benchmark Entrepreneurial Capital Demand Consider an entrepreneur in period j with assets a_j, genetic endowment g, pre-market type τ, and entrepreneurial productivity shock $u_{E,j}$. Let $f_{E,j} = f_E(g, \tau, u_{E,j})$. In the benchmark economy, operating capital is subject to the credit constraint

$$k_j \leq \bar{k}^{bm}(a_j) \equiv (1 + \chi^{bm})a_j, \qquad \chi^{bm} = 0.22.$$

Absent this constraint, the entrepreneur's optimal capital choice solves

$$\nu P f_{E,j} k_j^{\nu-1} = r + \delta,$$

so unconstrained capital demand is

$$k_j^u(g, \tau, u_{E,j}) = \left(\frac{r + \delta}{\nu P f_{E,j}}\right)^{\frac{1}{\nu-1}}. \tag{E.1}$$

The benchmark capital choice is therefore

$$k_j^{bm} = \min\{k_j^u, \bar{k}^{bm}(a_j)\}.$$

VC Financing Sector The VC counterfactual keeps the benchmark credit constraint fixed at $\chi^{bm} = 0.22$, but introduces an equity-financing channel. A VC investor provides supplemental capital in exchange for a share of the business. Let

$$\ell_j = \left[k_j - \bar{k}^{bm}(a_j)\right]_+$$

denote the amount of capital supplied by the VC sector. In exchange, the VC receives an equity share $s_j \in [0,1]$ of entrepreneurial output $Pf_{E,j}k_j^\nu$.

For a given entrepreneur and capital choice k_j, a competitive VC sector offers the minimum equity share that breaks even. The VC zero-profit condition is

$$s_j P f_{E,j} k_j^\nu = (r + \delta + r_p)\ell_j, \tag{E.2}$$

where $r_p \geq 0$ is the excess return required by VC investors. Hence, when $\ell_j > 0$, the required equity share is

$$s_j(k_j; r_p) = \frac{(r + \delta + r_p)\left[k_j - \bar{k}^{bm}(a_j)\right]}{P f_{E,j} k_j^\nu}. \tag{E.3}$$

If $s_j(k_j; r_p) > 1$, the VC contract is infeasible.

Substituting the zero-profit equity share into the entrepreneur's payoff gives

$$\big(1 - s_j(k_j; r_p)\big) P f_{E,j} k_j^\nu - (r + \delta) \min\{k_j, \bar{k}^{bm}(a_j)\}.$$

For $k_j > \bar{k}^{bm}(a_j)$, this is equivalent to

$$P f_{E,j} k_j^\nu - (r + \delta)\bar{k}^{bm}(a_j) - (r + \delta + r_p)\left[k_j - \bar{k}^{bm}(a_j)\right].$$

Thus, although the VC contract is written as equity finance, the entrepreneur's optimal capital choice is characterized by the same marginal condition as a premium-cost financing wedge:

$$\nu P f_{E,j} k_j^{\nu-1} = r + \delta + r_p \quad \text{whenever VC financing is active.}$$

Define

$$k_j^p(g, \tau, u_{E,j}; r_p) = \left(\frac{r + \delta + r_p}{\nu P f_{E,j}}\right)^{\frac{1}{\nu-1}}. \tag{E.4}$$

The optimal capital choice in the VC economy is therefore

$$k_j^{VC}(g, \tau, u_{E,j}, a_j; r_p) = \begin{cases} k_j^u, & \text{if } k_j^u \leq \bar{k}^{bm}(a_j), \\ \bar{k}^{bm}(a_j), & \text{if } k_j^u > \bar{k}^{bm}(a_j) \text{ and } k_j^p \leq \bar{k}^{bm}(a_j), \\ k_j^p, & \text{if } k_j^p > \bar{k}^{bm}(a_j) \text{ and } s_j(k_j^p; r_p) \leq 1. \end{cases} \tag{E.5}$$

Comparability with Credit-Relaxation Counterfactuals The credit-relaxation counterfactual raises the collateral parameter from χ^{bm} to a higher value χ. To compare this policy with the VC counterfactual, we choose the VC premium so that both economies generate the same aggregate amount of external entrepreneurial financing in stationary equilibrium.

For a credit-relaxation economy with collateral parameter χ, let $m_\chi^*(S, u)$ denote the stationary distribution, and let $k_\chi(S, u)$ and $\omega_\chi(S, u)$ denote the associated capital and occupation policies.

Aggregate entrepreneurial borrowing is

$$\mathcal{B}^{CR}(\chi) = \int \mathbb{I}_{\{\omega_\chi(S,u)=E\}} \left[k_\chi(S,u) - a(S)\right]_+ dm^*_\chi(S,u). \tag{E.6}$$

In the VC economy, the benchmark collateral limit remains fixed at χ^{bm}. For a given premium r_p, let $m^*_{VC}(S,u;r_p)$, $k^{VC}(S,u;r_p)$, and $\omega_{VC}(S,u;r_p)$ denote the stationary distribution and policy functions. Aggregate entrepreneurial borrowing is

$$\mathcal{B}^{VC}(r_p) = \int \mathbb{I}_{\{\omega_{VC}(S,u;r_p)=E\}} \left[k^{VC}(S,u;r_p) - a(S)\right]_+ dm^*_{VC}(S,u;r_p). \tag{E.7}$$

For each credit-relaxation economy χ, we choose the VC premium $r^*_p(\chi)$ to satisfy

$$\mathcal{B}^{VC}\left(r^*_p(\chi)\right) = \mathcal{B}^{CR}(\chi). \tag{E.8}$$

This matching condition equalizes total external entrepreneurial financing across the two counterfactuals while allowing the allocation of capital across entrepreneurs to differ.

General Equilibrium Solution Both the credit-relaxation and VC economies are solved as stationary recursive competitive equilibria. For each value of χ or r_p, households reoptimize occupational choice, consumption, saving, transfers, and capital demand. Prices (w, r) clear the same labor and capital markets as in the benchmark model.

For a given VC premium r_p, equilibrium prices are solved by damped fixed-point iteration:

$$\left(\log w^{(s+1)}, r^{(s+1)}\right) = \omega\, \mathcal{F}\left(\log w^{(s)}, r^{(s)}; r_p\right) + (1-\omega)\left(\log w^{(s)}, r^{(s)}\right), \tag{E.9}$$

where $\mathcal{F}$ denotes one application of the model map: solve the household problem, simulate the stationary distribution, compute aggregate labor and capital demand, and update prices to reduce market-clearing residuals. Table E.3 summarizes the computational algorithm.

The matched value $r^*_p(\chi)$ is the excess return required by VC investors when they provide money in exchange for a share of the business. The zero-profit condition converts this equity claim into an equivalent marginal financing wedge, which makes the VC counterfactual comparable to the credit-relaxation counterfactual in aggregate financing volume. The allocation of capital differs, however. A relaxation of χ expands collateralized borrowing as a function of assets, whereas VC finance supplies supplemental capital only when entrepreneurial productivity is high enough to justify the required ownership claim. Thus, the two regimes can deliver the same aggregate financing while allocating capital differently across g, τ, $u_{E,j}$, and wealth.

Table E.3: VC Counterfactual Algorithm: Matching Aggregate Entrepreneurial Borrowing

Step	Description
	Outer loop: for each target credit-relaxation economy χ
1	Solve the credit-relaxation economy and compute $\mathcal{B}^{CR}(\chi)$ using equation (E.6).
2	Initialize a bisection bracket $[r_p^{lo}, r_p^{hi}]$ and warm-start prices from the benchmark equilibrium.
	Bisection loop
3	Set $r_p^{mid} = (r_p^{lo} + r_p^{hi})/2$.
4	Given r_p^{mid}, solve the VC economy in general equilibrium using equation (E.9).
5	Compute aggregate entrepreneurial borrowing $\mathcal{B}^{VC}(r_p^{mid})$ using equation (E.7).
6	If $\mathcal{B}^{VC}(r_p^{mid}) > \mathcal{B}^{CR}(\chi)$, raise the lower bound r_p^{lo}. Otherwise, lower the upper bound r_p^{hi}.
7	Stop when $\left\lvert \mathcal{B}^{VC}(r_p^{mid}) - \mathcal{B}^{CR}(\chi) \right\rvert / \mathcal{B}^{CR}(\chi)$ is below the tolerance.
	Output
8	Record $r_p^*(\chi)$, equilibrium prices, policy functions, and the stationary distribution of the matched VC economy.

E.3 Decomposing the Ability–Finance Correlation

This appendix derives the decomposition used in Figure E.2. Entrepreneurial ability is specified as

$$\ln f_{E,j} = \beta_{0,E} + \beta_{g,E} + \beta_{\tau,E} + u_{E,j}. \tag{E.10}$$

Let

$$q_j \equiv [k_j - a_j]_+$$

denote external entrepreneurial finance. The correlation between log entrepreneurial ability and external finance is

$$\operatorname{corr}(\ln f_{E,j}, q_j) = \frac{\operatorname{Cov}(\ln f_{E,j}, q_j)}{\sigma_{\ln f_E} \sigma_q}.$$

Since the constant term $\beta_{0,E}$ does not affect covariance, equation (E.10) implies

$$\operatorname{Cov}(\ln f_{E,j}, q_j) = \operatorname{Cov}\left(\beta_{g,E}, q_j\right) + \operatorname{Cov}\left(\beta_{\tau,E}, q_j\right) + \operatorname{Cov}(u_{E,j}, q_j).$$

Dividing by $\sigma_{\ln f_E}\sigma_q$ gives

$$\operatorname{corr}(\ln f_{E,j}, q_j) = C_g + C_\tau + C_u, \tag{E.11}$$

where

$$C_g = \frac{\operatorname{Cov}(\beta_{g,E}, q_j)}{\sigma_{\ln f_E}\sigma_q}, \qquad C_\tau = \frac{\operatorname{Cov}(\beta_{\tau,E}, q_j)}{\sigma_{\ln f_E}\sigma_q},$$

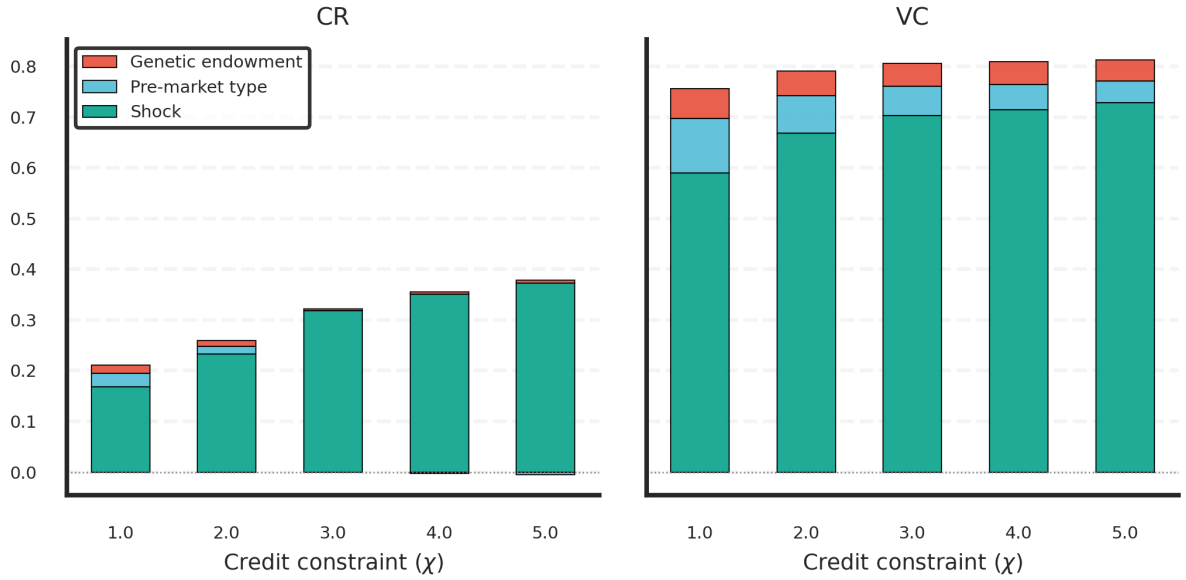
and

$$C_u = \frac{\operatorname{Cov}(u_{E,j}, q_j)}{\sigma_{\ln f_E}\sigma_q}.$$

Thus, C_g, C_τ, and C_u are component-specific covariance contributions to the overall ability-finance correlation. By construction, they sum exactly to $\operatorname{corr}(\ln f_{E,j}, q_j)$.

Figure E.2 reports this decomposition for the two financing regimes. The overall ability–finance correlation is much higher under VC than under broad credit relaxation. The sources also differ: under credit relaxation the correlation is increasingly driven by the transitory shock component C_u, while the persistent components C_g and C_τ shrink; under VC, capital continues to load on these persistent components.

Figure E.2: Decomposition of the Ability-Finance Correlation



Notes: The figure decomposes the correlation between log entrepreneurial ability $\ln f_{E,j}$ and external entrepreneurial finance $q_j = [k_j - a_j]_+$. The three stacked components correspond to the genetic component $\beta_{g,E}$, the pre-market type component $\beta_{\tau,E}$, and the shock $u_{E,j}$. The components are normalized covariance contributions, as derived in Appendix E.3. The left panel reports broad credit relaxation (CR), and the right panel reports venture-capital financing (VC). For each value of χ, the VC economy is matched to the CR economy by aggregate external entrepreneurial financing.