I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18113

The Heterosis Effect in Human Capital and Wealth Accumulation

Chen Zhu
Petri Böckerman

SEPTEMBER 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18113

The Heterosis Effect in Human Capital and Wealth Accumulation

Chen Zhu
China Agricultural University and GLO

Petri Böckerman
University of Jyväskylä, Research LABORE and IZA

SEPTEMBER 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Heterosis Effect in Human Capital and Wealth Accumulation

In genetics, heterosis describes how crossbreeding produces offspring with greater fitness than their parents. We propose a socioeconomic heterosis hypothesis: does genetic diversity at the individual level benefit economic success? Using UK Biobank data (N=488,152), we find that people with higher genome-wide heterozygosity perform better in modern societies. Greater heterozygosity is positively linked to education, income, leadership, height, and ownership of a home and car. A one standard deviation increase corresponds to about 0.75% higher income and modest gains in schooling and assets. Results are robust to additional controls and corrections for multiple testing, with no effects on migration, diabetes, or neuroticism. Effects rise steadily across the observed range and are stronger for men, suggesting sexual selection. Because heterozygosity is fixed at conception, our findings reveal an underappreciated endowment shaping human capital, wealth accumulation, and inequality.

JEL Classification: J10, J24, D31, I14

Keywords: heterosis, genetic heterozygosity, income, education, socioeconomic achievement, sexual selection

Corresponding author:
Petri Böckerman
Labour Institute for Economic Research
Pitkänsillanranta 3A
FI-00530 Helsinki
Finland

E-mail: petri.bockerman@labore.fi

Data availability

Eligible researchers may access UK Biobank data on www.ukbiobank.ac.uk, upon registration. For this study, permission to access the UK Biobank data was approved under the application numbered 89068.

Ethical approval

This study was conducted in accordance with the Declaration of Helsinki. Ethical approval for the UK Biobank was obtained from the NHS National Research Ethics Service (REC 11/NW/0382 and 16/NW/0274), granted in 2011 and renewed in 2016 and 2021. All participants provided written informed consent.

Funding

This research was supported by the National Natural Science Foundation of China (No. 72573164) and the 2115 Talent Development Program of China Agricultural University.

Competing interests

The authors declare no competing interests.

I. Introduction

Heterosis, also known as heterozygote advantage, refers to the genetic phenomenon whereby offspring resulting from genetically diverse parents (such as crossbreeding within a species between genetically distinct individuals) exhibit superior fitness or enhanced phenotypic characteristics (Chen, 2013; Lippman and Zamir, 2007). While extensive evidence of heterosis exists in both plants and animals, documentation of potential heterosis effects in humans beyond the molecular level remains limited. In this paper, we propose a novel socioeconomic heterosis hypothesis, suggesting that greater genetic diversity or individual heterozygosity may confer enhanced socioeconomic adaptability, resilience, and fitness within socioeconomic environments, reflected in improved cognitive/noncognitive abilities, socioeconomic achievements, fertility, and physical attributes.

Our socioeconomic heterosis hypothesis builds upon four distinct strands of biological and economic evidence. First, empirical findings have increasingly identified instances of molecular heterosis in humans (Comings and MacMurray, 2000; Aidoo et al., 2002). For example, Arnocky et al. (2021) reported that heterozygosity at immune-system loci (e.g., MHC genes) is associated with better immune response and developmental outcomes in men. Lewis (2010) proposed that enhanced physical attractiveness observed among mixed-race individuals might be explained by heterosis effects. Second, the well-documented phenomenon of inbreeding depression in human populations, characterized by adverse health outcomes from reduced genetic diversity (such as offspring of related individuals), indirectly implies a beneficial heterosis effect (Bittles, 2002; McQuillan et al., 2012; Yengo et al., 2021). Third, an emerging line of economic research shows that population-level genetic diversity or heterozygosity

correlates with economic development. Notably, Ashraf and Galor (2013) demonstrated that genetic diversity significantly correlates with regional economic development and innovation activity. Sequeira et al. (2019) showed that ancestral genetic diversity of population is beneficial to human capital development based on a cross-country analysis. Zhu et al. (2018) used geographic distances between parental hometowns as proxies for offspring heterosis and identified positive associations between these distances and offspring's educational attainment and height. Fourth, recent advances in genoeconomics have highlighted the role of genomic predisposition to certain traits—such as polygenic scores for educational attainment and Alzheimer's disease—in shaping human capital, wealth, and economic behavior (Barth et al., 2020; Shin et al., 2020; Papageorge and Thom, 2020; Rustichini et al., 2023; Gong et al., 2024; Houmark et al., 2024; Biroli et al., 2025). Despite these supportive findings, to our knowledge no prior study has empirically tested human heterosis effects using individual-level genetic data rather than indirect proxies.

In this study, we leverage the uniquely comprehensive UK Biobank dataset (N = 488,152) to directly investigate the heterosis effects in humans using individual-level genomic data. A key explanatory variable of individual-level heterosis is genome-wide heterozygosity. The measure is calculated as the proportion of autosomal single nucleotide polymorphism (SNPs) that are heterozygous (i.e., two different alleles present) using PLINK 2.0 (Purcell et al., 2007; Chang et al., 2015), and then adjusted for genetic principal components of ancestry (Bycroft et al., 2018). For each individual, we used a total of 605,876 autosomal SNPs to compute the heterozygosity score, covering all 22 autosomal chromosomes in human genome (i.e., genome-wide heterozygosity). Unlike prior studies that relied on population-level genetic distance or parental geographic separation as indirect proxies for heterosis (Ashraf and Galor, 2013;

Zhu et al., 2018), or that focused on polygenic scores derived from a limited number of SNPs, our approach offers a direct, precise, and individual-level measure of genome-wide genetic diversity. To our knowledge, this represents the first application of such a measure in the economic literature.

The *Darwinian fitness* originally refers to an organism's ability to survive and reproduce in natural environments, reflecting how well an organism is adapted to local conditions (Michod, 2000). In the socioeconomic context, researchers have operationalized this evolutionary concept using measurable socioeconomic outcomes, such as income, education, assets, fertility, and height (Nettle and Pollet, 2008; Borgerhoff Mulder and Beheim, 2011; Beauchamp 2016; Zhu et al., 2018). Building on this foundation, we expand the set of *socioeconomic fitness* indicators using UK Biobank data to include six core domains: (1) cognitive abilities (i.e., educational attainment, college degree); (2) non-cognitive abilities (i.e., neuroticism); (3) socioeconomic achievement (income, leadership positions, owning properties of house and car); (4) fertility (i.e., total number of children given birth or fathered); (5) physical characteristics (height, diabetes); and (6) economic behavior and preferences (i.e., migration and risk tolerance), resulting in a comprehensive set of 12 socioeconomic fitness outcomes. These measures collectively provide a comprehensive and multidimensional framework to empirically test the socioeconomic heterosis hypothesis in human populations.

It is worth noting that an individual's level of heterozygosity is established at the point of gamete formation *in utero* and is therefore predetermined, independent of any socioeconomic and confounding behavioral factors occurring after birth. Consequently, the key explanatory variable of genome-wide heterozygosity score in

our analysis does not suffer from endogeneity concerns due to reverse causality or selection bias, and can thus provide a "clean" estimate of heterosis effect. We thereby employ the high-dimensional fixed effects (HDFE) model to explicitly control for over 200 fixed effects and further support the robustness of our results.

Our empirical findings provide strong evidence for the existence of a heterosis effect in humans. Consistent with the predictions of the socioeconomic heterosis hypothesis, a one-standard-deviation increase in genome-wide heterozygosity is associated with a 0.75% increase in income, an additional 0.036 years of schooling, 0.0048 more offspring, a 0.082 cm increase in height, and a 0.002 increase in risk-tolerance score. Heterozygosity is also linked to higher probabilities of obtaining a college degree (0.26%), holding a leadership position (0.14%), owning a house (0.39%), and owning a car (0.66%). We control for age, sex, the top ten genetic principal components, and extensive regional and ethnic fixed effects in all models. The heterosis effects persist even after controlling for a variety of polygenic scores of cognitive and non-cognitive traits, indicating a robust heterosis effect directly resulting from genetic heterozygosity. Similar patterns are observed across both sexes, although the heterosis effect is notably stronger among males, potentially reflecting underlying sexual selection processes in socioeconomic environments. We find no significant associations between heterozygosity and migration behavior, diabetes risk, or neuroticism. Nonlinearity tests suggest that the association between heterozygosity and socioeconomic performance is monotonically increasing over the observed support, consistent with a private-benefit channel at the individual level.

The remainder of this paper is structured as follows. Section II introduces the theoretical foundations of the socioeconomic heterosis hypothesis. Section III

summarizes the data, describes variables used in the analysis, and details our empirical strategy. Section IV presents estimation results and supplemental robustness checks. Section V explores whether there is a hump-shaped association between genome-wide heterozygosity and socioeconomic outcomes at the individual level. Section VI concludes and points out directions for future research.

II. Theoretical Foundations of the Socioeconomic Heterosis Hypothesis

In evolutionary theory, Darwinian fitness describes an organism's ability to survive and reproduce in its environment, thereby passing on its genes to the next generation. It reflects how well adapted an organism is to local conditions and encompasses both survival and reproductive success (Michod, 2000). In ecology and biology, fitness is often context-dependent; traits advantageous in one environment may be disadvantageous in another. Notably, Darwin himself observed that introducing genetic variety (for example, cross-fertilizing plants from different lineages) often produced more vigorous offspring than self-fertilization. This early insight foreshadowed the concept of heterosis, wherein greater genetic diversity in offspring can lead to superior performance or resilience compared to their parents.

Translating these evolutionary ideas to humans, researchers in social sciences have asked what constitutes "fitness" in a modern socioeconomic context. In human societies, fitness can be thought of as the capacity to thrive and propagate under prevailing social and environmental conditions. Since direct measures of reproductive

success or survival can be complex in contemporary populations, scholars often use socioeconomic achievements as proxies for evolutionary success. For example, studies have operationalized human fitness using measurable outcomes that reflect success or status in society, such as: (1) educational attainment – education can enhance an individual's skills and opportunities, potentially influencing long-term family success; (2) income and wealth levels – higher economic resources may improve health and the ability to support more offspring; (3) asset ownership and material resources – possessing land, property, or other assets can improve security and multi-generational well-being; (4) reproductive success (fertility) – the number of children or timing of births directly ties into biological fitness in evolutionary terms; (5) health and stature – physical well-being and indicators such as height reflect nutrition and health during development, which can affect one's attractiveness or capability in social contexts.

Researchers have found that many of these factors correlate with an individual's reproductive outcomes or overall biological success in various populations (Nettle & Pollet, 2008; Borgerhoff Mulder & Beheim, 2011; Beauchamp, 2016; Zhu et al., 2018). In essence, higher socioeconomic performance – whether through wealth, education, or health – is often linked to greater Darwinian fitness in modern contexts. This approach links biology and economics by viewing prosperity and well-being as traits that may confer evolutionary advantages, such as improved survival or higher reproductive success within a given society.

Evolutionary principles have also been applied at the broader societal and historical level. A striking example is the macroeconomic evidence provided by Ashraf and Galor (2013), who examined how population- and country-level genetic diversity relates to long-run economic development. Their analysis of global populations found

a hump-shaped relationship between genetic diversity and economic performance across regions. Populations with intermediate genetic diversity historically achieved the highest levels of development and technological progress, whereas populations with very low diversity or extremely high diversity tended to have lower economic outcomes (Ashraf & Galor, 2013). The reasoning is that genetic diversity entails a trade-off between costs and benefits. On one hand, greater diversity can generate a broader range of talents, ideas, and innovations – analogous to the creative benefits of heterosis that enhance group problem-solving and adaptability. On the other hand, excessively high diversity might lead to communication barriers or reduce social cohesion, incurring social and coordination costs in a society. Ashraf and Galor's findings suggest that the most prosperous societies achieve a balance: sufficient genetic heterogeneity to foster innovation and adaptability, while avoiding the drawbacks of excessive fragmentation. This equilibrium parallels biological heterosis at the population level—a moderate degree of mixing that produces net positive effects on overall performance.

Building on these evolutionary insights, our socioeconomic heterosis hypothesis proposes that genetic diversity and mixing (outbreeding) can directly enhance individual performance and socioeconomic success. By analogy, the socioeconomic heterosis hypothesis suggests that individuals or populations with more diverse genetic backgrounds may exhibit advantages in traits that matter for success in modern societies. These advantages could include better health, higher cognitive ability, or greater creative potential – attributes that often translate into improved educational achievements, productivity, and other performance metrics. In other words, just as crossbred plants or animals can outperform their inbred counterparts, human groups that incorporate a healthy level of genetic variety might achieve higher performance in aggregate. This hypothesis creates a bridge between evolutionary biology and social

sciences. It implies that the evolutionary benefits of diversity (enhanced adaptability, resilience, and innovation capacity) manifest in measurable social and economic outcomes.

III. Data and empirical strategy

The main proprietary dataset we use is from the UK Biobank. UK Biobank is the database for a population-based study involving 502,409 UK residents approved by the NHS National Research Ethics Service (REC 11/NW/0382 and 16/NW/0274). Between March 2006 and July 2010, individuals residing within 25 miles of one of the 22 study assessment centers in England, Scotland, and Wales were recruited to provide data on a wide range of socio-demographic, clinical, and lifestyle outcomes. Blood, urine, and saliva samples, as well as physical measurements, were collected from all participants with their written informed consent during the interviews. For this study, permission to access and analyse the UK Biobank data was approved under the application 89068. The UK Biobank collects genetic information through the analysis of genotypes using the Affymetrix UK Biobank Axiom Array and the UK BiLEVE Axiom Array. The genetic data undergo quality control procedures, imputation, and principal components analysis, following the guidelines provided by Bycroft et al. (2018). The full UKB study protocol is available at https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf. Excluding individuals who did not have full information leads to an analytical sample of 488,152 individuals from the original UKB dataset. Table 1 provides summary statistics of the analytic sample.

Table 1. Summary statistics of the analytic sample

Variable	Mean/Percentage	Std. Dev.	Min	Max
ln(income)	10.69	0.69	9.80	12.21
Age	56.53	8.09	37	73
Male	45.6%	-	0	1
College	32.1%	-	0	1
Years of schooling	14.91	5.12	7	20
Leadership position	9.0%	-	0	1
Ownership of house	88.1%	-	0	1
Ownership of car	91.2%	-	0	1
Number of children	1.80	1.25	0	20
Neuroticism score	4.12	3.27	0	12
Distance migrated (in 1,000 km)	0.07	0.11	0.00	1.04
Risk taking	0.13	0.34	0	1
Height (in cm)	168.45	9.42	125	209
Diabetes	10.6%	-	0	1
Genetic variables:				
Heterozygosity score	0.190	0.002	0.124	0.222
Genetic PCA1	-1.13	54.28	-19.27	419.40
Genetic PCA2	0.36	27.86	-282.32	86.11
Genetic PCA3	-0.09	14.89	-143.26	98.82
Genetic PCA4	0.13	10.40	-107.57	37.10
Genetic PCA5	0.06	7.57	-21.96	36.22
Genetic PCA6	-0.05	5.01	-26.84	277.16
Genetic PCA7	0.00	4.98	-54.51	55.23
Genetic PCA8	-0.05	4.82	-135.84	53.84
Genetic PCA9	0.04	4.42	-36.58	12.61
Genetic PCA10	0.02	4.14	-54.15	44.57
Regions:				
East Midlands	6.5%	-	0	1
East of England	0.4%	-	0	1
London	28.3%	-	0	1
North East	11.1%	-	0	1
North West	11.6%	-	0	1
Scotland	0.1%	-	0	1
South East	4.3%	-	0	1
South West	6.3%	-	0	1
Wales	2.1%	-	0	1
West Midlands	14.6%	-	0	1
Yorkshire and The Humber	14.8%	-	0	1
N	488,152			

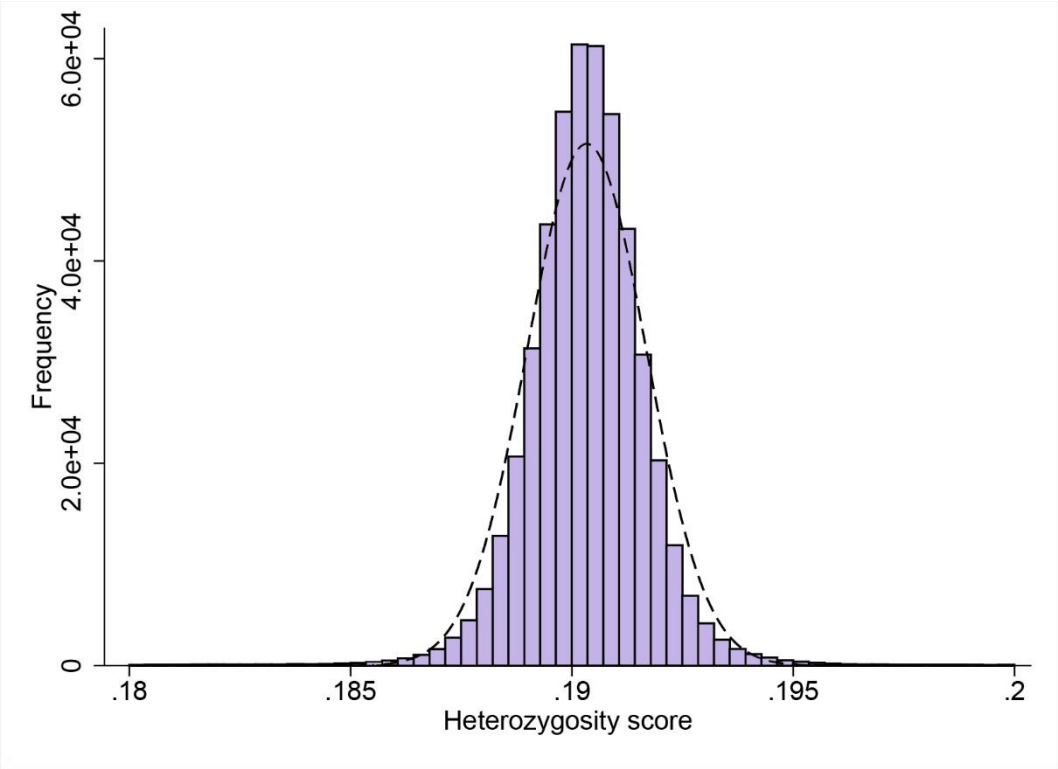
3.1. Measuring individual-level heterosis

We quantify individual-level heterosis using a genome-wide heterozygosity score. This

score is calculated as the proportion of autosomal SNPs that are heterozygous across all genotyped or imputed SNPs using PLINK 2.0. PLINK 2.0 is a widely used and state-of-the-art toolset for large-scale whole-genome association analyses (Purcell et al., 2007; Chang et al., 2015). To account for population structure, the raw heterozygosity score is adjusted for the top six genetic principal components of ancestry, following Bycroft et al. (2018).

Single nucleotide polymorphisms, or SNPs, are the most common form of genetic variation in the human genome (Syvänen, 2001). Each SNP represents a single base-pair difference in DNA sequence among individuals, and collectively, they serve as markers for genetic diversity, inheritance patterns, and disease risk. The presence of two different alleles at a SNP site indicates heterozygosity at that location. For each individual, the heterozygosity score is calculated from 605,876 autosomal SNPs spanning all 22 autosomal chromosomes in the human genome. Figure 1 presents the distribution of heterozygosity in our analytic sample, with a mean of 0.1903 and a standard deviation of 0.0017.

Figure 1. Distribution of the genome-wide heterozygosity score

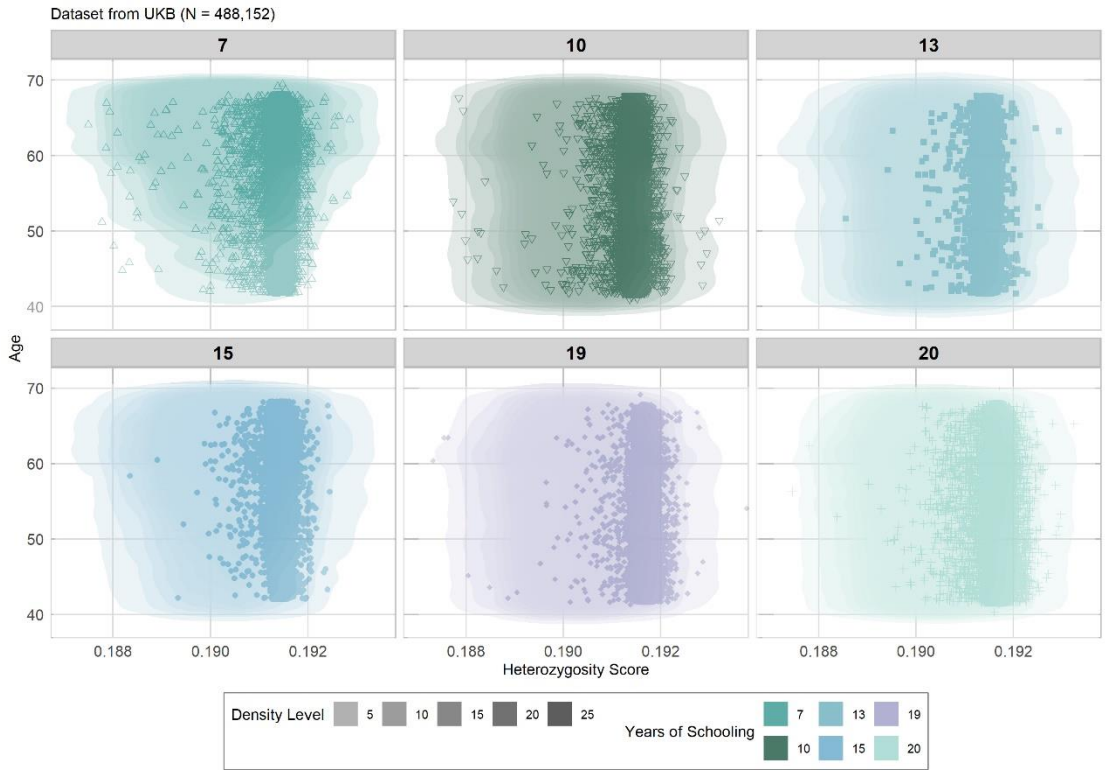


An important advantage of using the genome-wide heterozygosity score is that it provides a direct, individual-level measure of genetic diversity, rather than relying on indirect proxies such as population-level or country-level genetic distance or parental geographic separation. It captures the overall genetic variability across the entire genome, offering a more comprehensive and biologically grounded measure of potential heterosis. Moreover, because heterozygosity is fixed at conception and unaffected by subsequent environmental factors or individual choices, it provides as an exogenous and unbiased predictor in economic models, mitigating concerns about endogeneity and omitted variable bias.

To illustrate, Figure 2 displays the density distributions of heterozygosity scores by age and across different years of schooling, with age on the y-axis and

heterozygosity on the x-axis. Each panel/facet corresponds to a discrete level of educational attainment (i.e., 7, 10, 13, 15, 19, 20 years of schooling), with overlaid scatterplots and density contours. Across all panels, age distributions are generally consistent. Panels for 19 and 20 years of schooling show a slight rightward shift in the heterozygosity distribution, suggesting a potentially higher heterozygosity among individuals with the highest education levels.

Figure 2. Density distribution of the heterozygosity score by age and years of schooling



3.2. Measuring fitness and performance

To empirically evaluate the socioeconomic heterosis hypothesis, we employed twelve

fitness indicators from six key domains, reflecting a broad and multidimensional conception of human adaptability and success in socioeconomic environments:

(1) Cognitive abilities:

We included two core indicators: educational attainment (measured in years of schooling) and whether the individual has obtained a college degree. These measures serve as widely accepted proxies for cognitive skills and learning capacity, which are crucial determinants of human capital formation, labor market performance, and long-term socioeconomic outcomes.

(2) Non-Cognitive abilities:

We use the neuroticism score as a well-established measure of non-cognitive ability, capturing emotional stability and resilience to stress. In the UK Biobank, neuroticism is assessed using the 12-item scale from the Eysenck Personality Questionnaire-Revised Short Form (EPQ-R-S) (Smith et al., 2016). Higher scores indicate greater emotional reactivity and vulnerability to psychological distress, and have been consistently linked to poorer mental health, lower life satisfaction, and reduced productivity. Including this measure enables us to explore whether greater genetic diversity is associated with improved psychological resilience, thereby complementing our cognitive-based indicators of individual fitness.

(3) Socioeconomic achievement:

We use four indicators: individual income (log-transformed), occupying a leadership position (defined by whether an individual's occupation requires managing subordinates; Song et al., 2022), owning a house, and owning a car. These measures reflect career advancement and wealth accumulation. Leadership roles can serve as

signals of social influence, competence, and status within hierarchical structures.

(4) Fertility:

We include the number of biological children an individual has given birth to (for females) or fathered (for males) as a direct measure of reproductive fitness, which is a central concept in evolutionary theory.

(5) Physical characteristics:

We incorporate height and type 2 diabetes diagnosis to capture physical health and biological development. Height is commonly viewed as a cumulative marker of childhood nutrition and health, whereas type 2 diabetes serves as a proxy for metabolic risk. Both measures are closely linked to life-course productivity and longevity.

(6) Economic behavior and preferences:

Finally, we assess risk taking (defined by whether an individual selects "*Yes*" to the question "*Would you describe yourself as someone who takes risks?*") and migration behavior (defined as the geographic distance between an individual's current residence birthplace). Risk tolerance captures behavioral traits that shape socioeconomic choices such as entrepreneurship and investment, while migration reflects spatial mobility and adaptability to changing socioeconomic contexts.

Together, these twelve indicators enable a comprehensive, theory-informed analysis of how genome-wide heterozygosity may shape outcomes ranging from biological health to economic behavior. By addressing both evolutionary fitness (e.g., fertility, health) and modern economic outcomes (e.g., income, education, preferences), our approach places the socioeconomic heterosis hypothesis within an integrative framework connecting biology, economics, and the social sciences.

3.3 Other covariates

All models include a comprehensive set of covariates to ensure that our estimates of the heterosis effect are not confounded by demographic, population-structural, environmental, or genetic factors. Specifically, we control for age and sex (coded as male = 1, female = 0), as well as the top ten genetic principal components of ancestry (i.e., genetic PCAs). These genetic PCAs are derived from genome-wide SNP data and account for subtle population structure and ancestral background, thereby mitigating potential bias arising from population stratification.

To further account for spatial heterogeneity, we include a full set of regional fixed effects. Using the east and north coordinates of participants' residential addresses, we construct dummy variables for 11 broad UK regions: East Midlands, East of England, London, North East, North West, Scotland, South East, South West, Wales, West Midlands, and Yorkshire and The Humber. These regional indicators help to account for unobserved, time-invariant environmental and institutional characteristics that may systematically vary across geographic areas—such as labor market conditions, public service availability, healthcare infrastructure, and historical development patterns.

Moreover, we control for ethnicity using detailed self-reported ethnic background data provided by UKB participants. Ethnicity dummies allow us to further adjust for sociocultural and environmental differences that may correlate with both heterozygosity and the outcomes of interest.

By incorporating both regional and ethnicity fixed effects, as well as their interactions in some specifications, we flexibly account for a wide range of contextual

factors. This approach ensures that our results are not driven by latent regional or ethnic clustering in genetic variation or outcome distributions, and allows us to isolate the contribution of genome-wide heterozygosity to individual-level performance more precisely and robustly.

3.4 Empirical strategy

In estimation, we restrict the sample to participants with complete information on all key dependent and independent variables (N = 488,152). We employ a high-dimensional fixed effects model that allows us to include multiple fixed effects (Guimaraes and Portugal, 2010). We estimate HDFE models with the following structure:

$$Y = \lambda\mathcal{H} + X\beta + \theta_1 Region + \theta_2 Ethnic\_group + +\theta_3 Region \times Ethnic\_group + \varepsilon \quad (1)$$

where Y contains participants' twelve measures of fitness and performance as discussed in section 3.2. $\mathcal{H}$ represents the genome-wide heterozygosity as discussed in section 3.1, and λ is the coefficient of our primary interest. X contains a set of observed individual socio-demographic characteristics, including age, sex, and top 10 genetic PCAs. We also include a total of 234 fixed effects in the estimation: 10 region FEs, 22 ethnic group FEs, and 202 region × ethnic group FEs. ε is a stochastic error term assumed to be distributed $N(0, \sigma^2)$. To minimize the Type I errors, we additionally apply the Bonferroni correction to account for multiple comparisons across the 12 outcome variables, i.e., Bonferroni corrected significance level = 0.05/12 = 0.00417.

IV. Estimation results

4.1 The effects of heterozygosity on fitness and performance

Table 2 presents the results of high-dimensional fixed effects estimations for the benchmark equation 1, examining the association between genome-wide heterozygosity and twelve indicators of socioeconomic fitness and performance. Panel A reports coefficient estimates and standard errors (in parentheses), while Panel B shows corresponding p-values and Bonferroni-adjusted significance levels for the key explanatory variable of heterozygosity score. When interpreting the regression results, it is important to note that unobserved heterogeneity across regions, ethnicities, and their interactions is accounted for through the extensive set of fixed effects described in Section III.

The results in Panel A show that genome-wide heterozygosity (Ψ) is positively and significantly associated with a range of outcomes, including educational attainment (columns 1–2), socioeconomic achievement (columns 4–7), fertility (column 8), height (column 9), and risk tolerance (column 11). These associations remain intact after accounting for age, sex, and top ten genetic PCAs. As expected, male and age are also significantly linked to several outcomes such as income, education, and physical characteristics. Panel B shows that all associations remain statistically significant after Bonferroni correction, except for number of children (column 8), which becomes marginally insignificant under the most conservative adjustment for multiple comparisons.

Quantitatively, a one-standard-deviation increase in heterozygosity corresponds to a 0.75% rise in income, an additional 0.036 years of schooling, a 0.082 cm increase

in height, a 0.002 increase in risk-tolerance score, and higher likelihoods of obtaining a college degree (0.26%), holding a leadership position (0.14%), owning a house (0.39%), and owning a car (0.66%). These results reveal that greater genetic heterozygosity enhances individual adaptability and performance across cognitive, physical, and economic domains in modern socioeconomic environments, which provide compelling empirical support for the “socioeconomic heterosis hypothesis”.

Table 2. HDFE estimation results of the heterosis effects with Bonferroni correction

	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）	（9）	（10）	（11）	（12）
	Years of schooling	College	Neuroticism score	ln(income)	Leadership position	Ownership of house	Ownership of car	Number of children	Height (in cm)	Diabetes	Risk taking	Distance migrated (in 1,000 km)
A.												
Heterozygosity score	0.0358***	0.0026***	-0.0074	0.0075***	0.0014***	0.0039***	0.0066***	0.0048***	0.0815***	-0.0003	0.0023***	-0.0003*
	(0.0073)	(0.0007)	(0.0054)	(0.0010)	(0.0004)	(0.0005)	(0.0012)	(0.0018)	(0.0092)	(0.0004)	(0.0005)	(0.0002)
Age	-0.1372***	-0.0070***	-0.0403***	-0.0288***	-0.0052***	0.0022***	-0.0136***	0.0263***	-0.1580***	0.0018***	-0.0068***	0.0001***
	(0.0009)	(0.0001)	(0.0006)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0011)	(0.0000)	(0.0001)	(0.0000)
Male	0.7618***	0.0311***	-0.9525***	0.1205***	0.0568***	-0.0175***	0.0696***	-0.0407***	13.2741***	0.0211***	0.1294***	0.0007*
	(0.0143)	(0.0013)	(0.0104)	(0.0020)	(0.0008)	(0.0009)	(0.0024)	(0.0035)	(0.0182)	(0.0008)	(0.0009)	(0.0003)
Top 10 Genetic PCAs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnic group FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region*Ethnic group FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Total No. of FE	234	234	234	234	234	234	234	234	234	234	234	234
Observations	488,152	488,152	488,152	488,152	488,152	488,152	488,152	488,152	488,152	488,152	488,152	488,152
B.												
p-value of Heterozygosity score	8.60E-07	1.17E-04	1.71E-01	3.40E-13	7.06E-04	1.96E-17	7.76E-08	6.42E-03	8.62E-19	4.04E-01	1.70E-06	7.67E-02
Significant after Bonferroni correction ($p < 0.05/12 = 0.00417$)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

4.2 Gender differences in heterosis effects and the role of sexual selection

To explore potential gender-specific patterns in the heterosis effect, we conducted separate high-dimensional fixed effects estimations for male and female subsamples. Table 3 presents the results, with Panel A for males and Panel B for females, including Bonferroni correction results.

Three important patterns can be detected. First, in males, genome-wide heterozygosity is positively and significantly associated with 8 out of 12 socioeconomic outcomes after Bonferroni correction, including years of schooling (column 1), college attainment (column 2), income (column 4), leadership roles (column 5), home and car ownership (column 6 and 7), height (column 9), and risk tolerance (column 11). Second, for females, significant associations are only limited to income, home ownership, and height (3 out of 12 outcomes) after Bonferroni correction. Third, the magnitude of these effects is consistently greater in males. To illustrate this pattern, a one-standard-deviation increase in heterozygosity corresponds to a 0.92% rise in income for males, compared to a 0.57% rise for females (column 4). Similarly, the increase in height is 0.099 cm for males and 0.064 cm for females (column 9).

These gender disparities align with evolutionary theories of *sexual selection*, which describe how certain traits increase reproductive success by enhancing an individual's attractiveness to potential mates or competitiveness in securing them. In many mammalian species, males exhibit exaggerated traits (such as larger body size) driven by both inter-sexual selection (female choice) and intra-sexual selection (male-male competition) (Janicke and Fromonteil, 2021). This asymmetry arises because males typically compete for mating opportunities, while females are more selective in choosing partners. Comparable dynamics appear to operate in human societies, where

traits such as height, income, and leadership status function as signals of male quality and influence outcomes in the marriage market (Zhu et al., 2018). These attributes, however, are generally less consequential for female success, which may explain the stronger heterosis effects observed in males and the attenuated effects in females.

To conclude, these findings suggest that sexual selection and intra-sexual competition continue to operate in modern socioeconomic environments. They lend further support to the socioeconomic heterosis hypothesis by illustrating how genetic diversity interacts with pressures of society to shape differential outcomes across gender in human capital and socioeconomic achievement.

Table 3. HDFE estimation results by genders

	(1) Years of schooling	(2) College	(3) Neuroticism score	(4) ln(income)	(5) Leadership position	(6) Ownership of house	(7) Ownership of car	(8) Number of children	(9) Height (in cm)	(10) Diabetes	(11) Risk taking	(12) Distance migrated (in 1,000 km)
A. Males (N = 222,597)												
Heterozygosity score	0.0476***	0.0029***	-0.0128*	0.0092***	0.0024***	0.0042***	0.0087***	0.0060**	0.0993***	0.0000	0.0034***	-0.0003
	(0.0106)	(0.0010)	(0.0077)	(0.0015)	(0.0007)	(0.0007)	(0.0018)	(0.0026)	(0.0138)	(0.0006)	(0.0008)	(0.0003)
Age	-0.1108***	-0.0053***	-0.0417***	-0.0270***	-0.0063***	0.0025***	-0.0088***	0.0262***	-0.1566***	0.0022***	-0.0083***	0.0001***
	(0.0013)	(0.0001)	(0.0009)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0003)	(0.0017)	(0.0001)	(0.0001)	(0.0000)
p-value of Heterozygosity score	7.44E-06	2.92E-03	9.55E-02	4.71E-10	3.12E-04	2.46E-10	1.19E-06	2.38E-02	6.63E-13	9.82E-01	4.14E-05	2.65E-01
Significant after Bonferroni correction (p < 0.05/12 = 0.00417)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No
B. Females (N= 265,555)												
Heterozygosity score	0.0232**	0.0023**	-0.0020	0.0057***	0.0005	0.0033***	0.0044***	0.0045*	0.0638***	-0.0006	0.0011**	-0.0004
	(0.0100)	(0.0009)	(0.0076)	(0.0014)	(0.0005)	(0.0006)	(0.0017)	(0.0024)	(0.0123)	(0.0005)	(0.0005)	(0.0003)
Age	-0.1608***	-0.0085***	-0.0389***	-0.0305***	-0.0043***	0.0019***	-0.0179***	0.0264***	-0.1597***	0.0015***	-0.0055***	0.0001***
	(0.0012)	(0.0001)	(0.0009)	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0003)	(0.0015)	(0.0001)	(0.0001)	(0.0000)
p-value of Heterozygosity score	2.04E-02	1.10E-02	7.90E-01	7.36E-05	3.52E-01	1.10E-07	9.49E-03	5.71E-02	2.30E-07	2.72E-01	3.28E-02	1.60E-01
Significant after Bonferroni correction (p < 0.05/12 = 0.00417)	No	No	No	Yes	No	Yes	No	No	Yes	No	No	No

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

4.3 Adjustment for inherited genetic predispositions through polygenic scores

A key concern in interpreting the estimated heterosis effect is the potential for omitted variable bias due to unobserved genetic endowments or correlated genetic factors. Specifically, individuals with higher cognitive or non-cognitive potential—shaped by their inherited genetic traits—may be more likely to migrate or form partnerships across genetically diverse backgrounds. Consequently, their offspring would not only possess greater genetic heterozygosity through parental mixing but also inherit favourable genetic endowments for socioeconomic success. In this case, the observed heterosis effect may partly capture the direct influence of inherited endowments rather than heterozygosity itself.

To address this concern, we leverage the rich genomic and multi-omics data in the UK Biobank to construct polygenic scores that proxy for individual-level genetic predispositions to key behavioral, psychological, and physiological traits. Polygenic scores (PGSs), or polygenic indices (PGIs), have been used in economic analyses as measures of genetic endowments of complex socioeconomic traits (Barth et al., 2020; Papageorge and Thom, 2020; Rustichini et al., 2023; Houmark, et al., 2024; Biroli et al., 2025). These scores help isolate the heterosis effect from potential confounding arising from direct hereditary transmission. Specifically, we construct and control for polygenic scores for educational attainment (Okbay et al., 2022), neuroticism (Baselmans et al., 2019), externalizing behaviors (Linnér et al., 2021), and risk-taking tendency (Linnér et al., 2019), which are particularly relevant given their known influence on economic outcomes. Moreover, we incorporate PGSs for body mass index (BMI), height, risks of schizophrenia, hypertension, bipolar disorder, cardiovascular disease (CVD), and type 2 diabetes, which are directly available in UK Biobank and

may reflect health-related productivity.

Polygenic scores are constructed using PLINK 2.0, a state-of-the-art software tool for genome-wide association and prediction analysis (Purcell et al., 2007; Chang et al., 2015). Each PGS aggregates the weighted sum of alleles associated with a given trait based on large-scale genome-wide association studies (GWAS) summary statistics (Rustichini et al., 2023; Wang et al., 2020). By including these PGSs in our estimation, we substantially mitigate the risk of omitted variable bias stemming from inherited genetic advantages, allowing for a more rigorous and credible identification of the heterosis effect on socioeconomic outcomes.

As presented in Table 4, the inclusion of PGSs does not attenuate the estimated effects of genome-wide heterozygosity. In fact, some associations become more pronounced, providing stronger empirical support for the socioeconomic heterosis hypothesis. For instance, a one-standard-deviation increase in heterozygosity is after adjustment for PGSs associated with a 1.02% rise in income (up from 0.75%) and an additional 0.064 years of schooling (up from 0.036 years).

Several polygenic scores independently contribute to the socioeconomic outcomes, consistent with prior research on the role of genetic endowments in wealth accumulation (e.g., Barth et al., 2020). For example, the PGS of educational attainment is positively associated with multiple indicators of socioeconomic success (including income, educational level, leadership roles, and property ownership) but negatively associated with number of children. Conversely, higher genetic risk for schizophrenia is significantly associated with lower income, education, leadership likelihood, and asset ownership.

Overall, our results suggest that failing to control for inherited genetic

predispositions may lead to an underestimation of the heterosis effect. It also implies that, when examining the relationship between genetic predisposition and socioeconomic outcomes, it is important to distinguish between the effects of genetic diversity and specific genetic endowments.

Table 4. HDFE estimation results after adjustment for polygenic scores

	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）	（9）	（10）	（11）	（12）
	Years of schooling	College	Neuroticism score	ln(income)	Leadership position	Ownership of house	Ownership of car	Number of children	Height (in cm)	Diabetes	Risk taking	Distance migrated (in 1,000 km)
Heterozygosity score	0.0641***	0.0052***	-0.0121**	0.0102***	0.0014***	0.0045***	0.0073***	0.0042**	0.0971***	-0.0007*	0.0025***	-0.0000
	(0.0071)	(0.0006)	(0.0055)	(0.0010)	(0.0004)	(0.0005)	(0.0013)	(0.0018)	(0.0082)	(0.0004)	(0.0005)	(0.0002)
Age	-0.1406***	-0.0073***	-0.0399***	-0.0291***	-0.0052***	0.0021***	-0.0137***	0.0264***	-0.1609***	0.0019***	-0.0068***	0.0001***
	(0.0009)	(0.0001)	(0.0006)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0010)	(0.0000)	(0.0001)	(0.0000)
Male	0.8008***	0.0346***	-0.9541***	0.1243***	0.0570***	-0.0169***	0.0701***	-0.0406***	13.2997***	0.0207***	0.1299***	0.0010***
	(0.0137)	(0.0013)	(0.0103)	(0.0020)	(0.0008)	(0.0009)	(0.0024)	(0.0035)	(0.0159)	(0.0008)	(0.0009)	(0.0003)
PGS of educational attainment	14.2086***	1.3278***	-1.4486***	0.9600***	0.0396***	0.2346***	0.2215***	-0.4752***	3.0228***	-0.0128***	0.1548***	0.1320***
	(0.0657)	(0.0060)	(0.0496)	(0.0095)	(0.0039)	(0.0043)	(0.0118)	(0.0169)	(0.0763)	(0.0038)	(0.0045)	(0.0016)
PGS of neuroticism	-1.4036***	-0.1291***	1.9616***	-0.0321***	0.0030	-0.0036	0.0362***	0.0333**	-0.5120***	-0.0006	0.0058	-0.0122***
	(0.0626)	(0.0057)	(0.0473)	(0.0090)	(0.0037)	(0.0041)	(0.0112)	(0.0161)	(0.0726)	(0.0037)	(0.0043)	(0.0015)
PGS of externalizing behaviors	0.0224	0.0029	-0.2471***	0.0202*	0.0029	-0.0044	0.0118	0.0102	-0.0586	0.0050	0.0090*	0.0058***
	(0.0736)	(0.0068)	(0.0556)	(0.0106)	(0.0044)	(0.0048)	(0.0132)	(0.0189)	(0.0854)	(0.0043)	(0.0051)	(0.0018)
PGS of risk-taking tendency	-0.9423***	-0.0918***	0.0840*	-0.0450***	0.0118***	-0.0474***	-0.0291***	0.0779***	-0.3380***	-0.0024	0.0644***	-0.0082***
	(0.0622)	(0.0057)	(0.0471)	(0.0089)	(0.0037)	(0.0041)	(0.0111)	(0.0160)	(0.0722)	(0.0036)	(0.0043)	(0.0015)
PGS of body mass index	-0.0196***	-0.0034***	-0.0270***	-0.0076***	0.0008*	-0.0044***	-0.0028**	0.0091***	0.0137*	0.0028***	0.0018***	-0.0005**
	(0.0072)	(0.0007)	(0.0054)	(0.0010)	(0.0004)	(0.0005)	(0.0013)	(0.0018)	(0.0083)	(0.0004)	(0.0005)	(0.0002)
PGS of height	-0.0037	0.0007	-0.0119	0.0184***	0.0026***	0.0036***	0.0136***	0.0013	4.5256***	-0.0002	0.0048***	0.0007***
	(0.0102)	(0.0009)	(0.0078)	(0.0015)	(0.0006)	(0.0007)	(0.0018)	(0.0026)	(0.0119)	(0.0006)	(0.0007)	(0.0003)
PGS of type 2 diabetes	-0.0219***	-0.0027***	0.0223***	-0.0054***	-0.0006	-0.0046***	-0.0058***	-0.0020	-0.0117	0.0253***	-0.0018***	-0.0003*
	(0.0075)	(0.0007)	(0.0056)	(0.0011)	(0.0004)	(0.0005)	(0.0013)	(0.0019)	(0.0086)	(0.0004)	(0.0005)	(0.0002)
PGS of schizophrenia	-0.0517***	-0.0016**	0.1077***	-0.0202***	-0.0046***	-0.0092***	-0.0268***	-0.0001	0.0183**	-0.0026***	-0.0069***	-0.0017***
	(0.0074)	(0.0007)	(0.0056)	(0.0011)	(0.0004)	(0.0005)	(0.0013)	(0.0019)	(0.0086)	(0.0004)	(0.0005)	(0.0002)
	0.0686***	0.0083***	0.0425***	0.0068***	0.0009**	0.0005	0.0018	0.0085***	-0.0232***	0.0000	0.0018***	0.0013***

PGS of bipolar disorder	(0.0071)	(0.0007)	(0.0054)	(0.0010)	(0.0004)	(0.0005)	(0.0013)	(0.0018)	(0.0082)	(0.0004)	(0.0005)	(0.0002)
PGS of hypertension	-0.0552***	-0.0049***	0.0583***	-0.0086***	-0.0009*	-0.0027***	-0.0054***	0.0006	-0.0455***	0.0012***	-0.0027***	-0.0011***
	(0.0077)	(0.0007)	(0.0058)	(0.0011)	(0.0005)	(0.0005)	(0.0014)	(0.0020)	(0.0089)	(0.0004)	(0.0005)	(0.0002)
PGS of CVD	-0.0142*	-0.0012*	0.0018	-0.0023**	0.0001	-0.0012**	0.0001	0.0056***	-0.1039***	0.0006	0.0002	-0.0002
	(0.0073)	(0.0007)	(0.0055)	(0.0011)	(0.0004)	(0.0005)	(0.0013)	(0.0019)	(0.0085)	(0.0004)	(0.0005)	(0.0002)
Top 10 Genetic PCAs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnic group FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region*Ethnic group FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value of Heterozygosity score	1.22E-19	1.07E-15	2.71E-02	4.96E-23	8.17E-04	2.93E-22	7.64E-09	2.03E-02	1.34E-32	7.49E-02	1.90E-07	8.82E-01
Significant after Bonferroni correction ($p < 0.05/12 = 0.00417$)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No

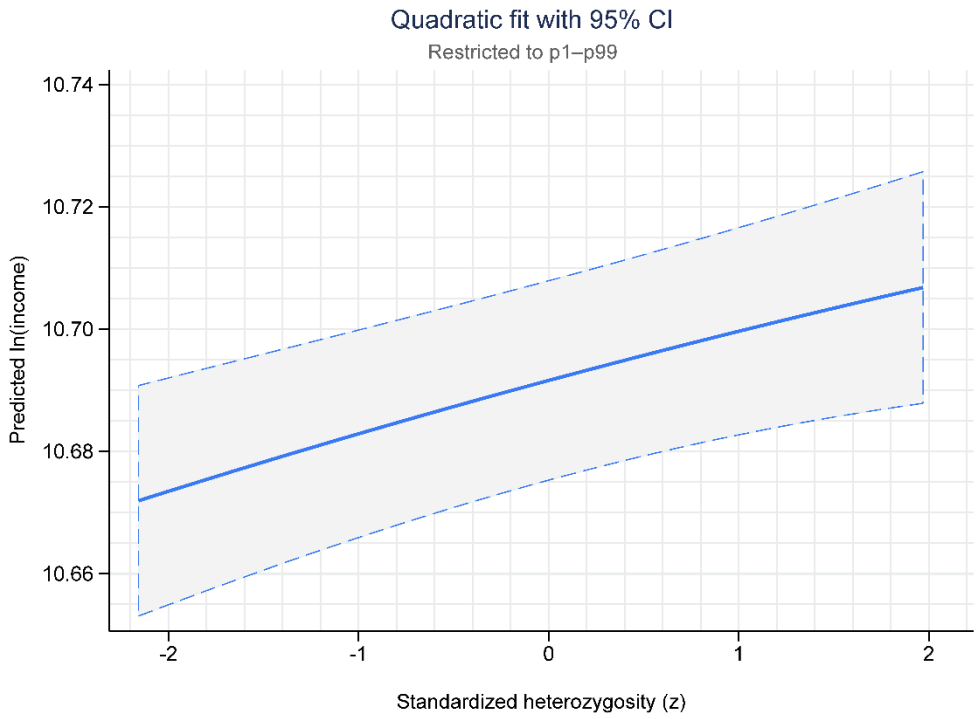
Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

V. Is there an optimal level of heterozygosity?

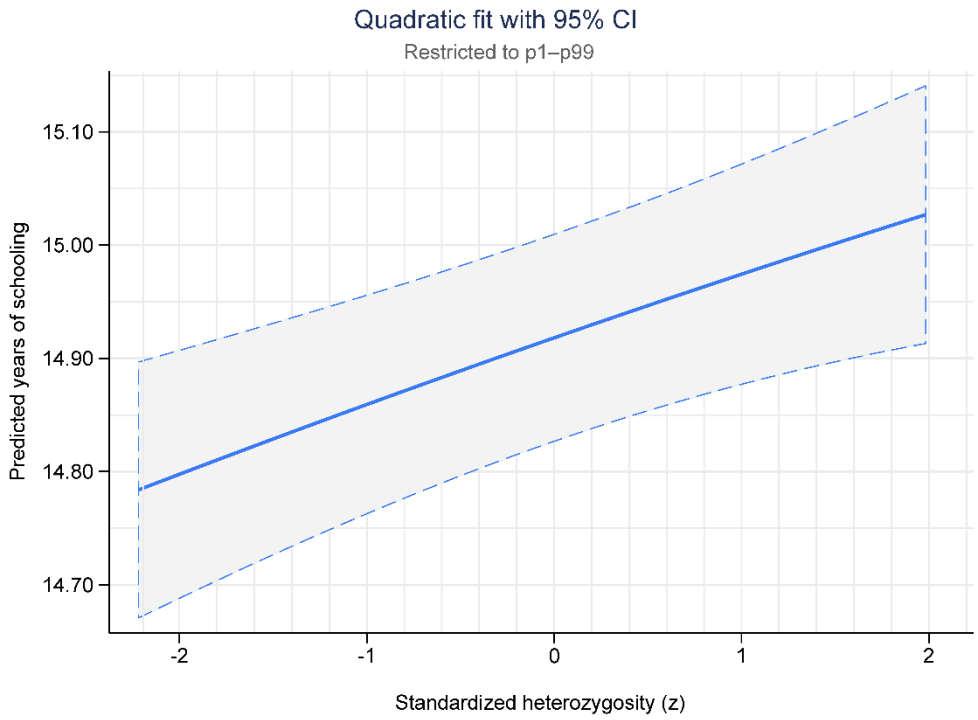
Ashraf and Galor (2013) document a hump-shaped association between population genetic diversity and long-run development. Whether a similar interior optimum exists at the individual level is ultimately an empirical question. We therefore test for an interior maximum in the relationship between genome-wide heterozygosity and socioeconomic outcomes using the quadratic U-shape procedure of Lind and Mehlum (2010). Specifically, we estimate a quadratic specification in heterozygosity with the full set of controls and fixed effects, and then evaluate the slope at the lower and upper bounds of the observed support (Figure 3). The Lind-Mehlum tests reject the existence of an interior optimum for neither log of income nor years of schooling in our data. For log of income, the slope at the 1st percentile of heterozygosity is positive (0.0099, $p < 0.001$), and the slope at the 99th percentile remains positive (0.0071, $p = 0.011$). Similarly, for years of schooling, the slope at the 1st percentile of heterozygosity is positive (0.0635, $p < 0.001$), and the slope at the 99th percentile remains positive (0.0522, $p = 0.006$). Thus, within the UK Biobank sample, the association between heterozygosity and socioeconomic performance is monotonically increasing over the empirically relevant range; any implied turning point from the quadratic lies beyond the observed support.

Figure 3. Predicted outcomes by standardized heterozygosity

A. ln(income)



B. Years of schooling



Notes: Curves show quadratic fit with 95% confidence intervals (CI) from regressions of ln(income) (panel A) and years of schooling (panel B) on standardized heterozygosity and its square, controlling for age, sex, polygenic scores, 10 ancestry PCs, ethnicity and region fixed effects; SEs clustered by location. X-axis restricted to the 1st–99th percentiles of standardized heterozygosity.

Why might a hump-shaped relationship emerge in country-level analyses but not at the individual level in our setting? First, the costs emphasized at the macroeconomic level (e.g., communication frictions, lower social cohesion, and coordination failures at high diversity) are largely group-level externalities; they need not be borne by an individual's own heterozygosity. By contrast, the benefits of heterozygosity (e.g., masking deleterious recessives, improved immunocompetence, stress robustness) accrue privately and can raise the certainty-equivalent return to human-capital investment and to sorting into high-variance occupations. This asymmetry naturally favours monotonic individual-level gains within a relatively homogeneous institutional environment. Second, the support of heterozygosity in our data is relatively narrow: UK Biobank contains predominantly individuals with broadly similar ancestry. A global hump-shaped relationship can still be consistent with a locally increasing segment when the sample does not reach the range where diminishing returns (if any) set in.

These findings should be interpreted with two caveats. First, extrapolation beyond the observed support is unwarranted: our tests speak to the UK Biobank range and do not rule out curvature in broader multi-ancestry or cross-country samples. Second, any "optimal" level is likely to be environment-dependent. If heterozygosity acts partly through robustness to environmental stressors (pollution, pathogen load, earnings volatility), the location of a peak may shift across contexts; in harsh environments, the marginal benefit of heterozygosity could remain positive over a

larger range.

To summarize, within the UK Biobank's ancestry composition and institutional context, we find no evidence of an interior optimum: socioeconomic returns to heterozygosity increase throughout the observed range. This micro-level monotonicity can coexist with the macro-level hump documented by Ashraf and Galor (2013) once one recognizes that the relevant costs arise primarily at the group level, whereas the benefits accrue to individuals—a distinction made explicit by our theory and evidence.

VI. Conclusion and discussion

This study provides empirical support for the socioeconomic heterosis hypothesis, indicating that greater genetic heterozygosity (genetic diversity within an individual's genome) enhances adaptability and performance across multiple domains in modern societies. Using the extensive UK Biobank dataset, we find that individuals with higher genome-wide heterozygosity tend to achieve better socioeconomic outcomes. Notably, higher genetic diversity is significantly associated with higher income, greater educational attainment, more frequent leadership roles, increased asset ownership, taller stature, and greater risk tolerance. These positive associations remain strong even after controlling for comprehensive polygenic scores related to cognitive ability and non-cognitive traits, suggesting that the heterosis effects are not merely artifacts of specific inherited genes for intelligence or personality. In other words, beyond specific genetic endowments, genetic diversity itself appears to confer broad advantages, facilitating greater human capital accumulation and success across diverse socioeconomic domains.

Our findings highlight the theoretical importance of genetic diversity as a factor in economic success. In the context of human capital theory, genetic heterozygosity can be viewed as an innate component of human capital. Greater genetic diversity may improve developmental robustness and cognitive or health outcomes, thereby improving an individual's capacity to learn, innovate, and work productively. This biological advantage complements traditional human capital inputs (like formal schooling and training) by enhancing the effectiveness with which individuals can acquire and use their skills. Empirically, the results are noteworthy because they introduce genetic heterogeneity as a novel predictor of economic outcomes, supported by large-scale data. The robustness of the heterosis effect across multiple outcomes and controls underscores that genetic diversity is a non-negligible determinant of socioeconomic achievement, meriting attention alongside more commonly studied factors such as education, family background, and environment.

Notably, the socioeconomic heterosis effects we observe are more pronounced in males than in females, a pattern that aligns with evolutionary theories of sexual selection and their socioeconomic manifestations. In many species, including humans, males have historically faced strong competitive pressures—both intersexual (attracting mates) and intrasexual (competing with other males)—leading to the exaggeration of certain traits. By contrast, women's socioeconomic outcomes tend to depend relatively more on other factors, which could explain why increased genetic diversity leads to slightly smaller measurable benefits for females in the UK biobank data. In essence, evolutionary pressures may have made male economic and physical traits more sensitive to the benefits of genetic diversity, resulting in a stronger heterosis effect among men.

Methodologically, a key feature of our analysis is the inclusion of polygenic scores to account for inherited abilities and trait endowments. This helps address potential omitted variable bias by accounting for known genetic influences on outcomes. Interestingly, we find that the estimated benefits of heterozygosity remain robust or even become larger after adjusting for these polygenic scores. For instance, a one-standard-deviation increase in genome-wide heterozygosity is associated with approximately a 1.0% increase in income after controlling for PGSs – up from about 0.75% when such genetic controls are absent. Likewise, the estimated effect of heterozygosity on educational attainment nearly doubles—though it remains modest in magnitude—once polygenic factors are accounted for.

These comparisons indicate that failing to control for specific genetic predispositions can lead to an underestimation of the heterosis effect. In other words, part of the heterosis advantage was initially obscured by genetic factors correlated with both diversity and achievement. Once those factors are held constant, the distinct contribution of genetic diversity becomes evident. This insight is empirically important because it differentiates the impact of overall genetic diversity from the influence of particular advantageous genes: our results indicate that it is the combination of diverse genes, rather than any single gene, that gives individuals an edge in socioeconomic outcomes.

From a broader economics perspective, our individual-level evidence offers a novel lens on the ongoing debate about genetics and economic development. There is an emerging body of macroeconomic research suggesting that genetic diversity within populations can impact innovation, institutional dynamics, and long-run growth prospects. For example, diversity may foster a wider range of ideas and talents, which

can boost creativity and productivity, though extremely high diversity could also pose coordination challenges (Ashraf and Galor, 2013). Our findings contribute a crucial micro-foundation to this discourse: at the individual level, genetic heterogeneity confers advantages that likely aggregate to meaningful economic gains for society. By demonstrating that genetically diverse individuals tend to have higher human capital and wealth accumulation, we provide evidence consistent with the notion that moderate genetic diversity is economically beneficial. This linkage between biology and economics enriches theories of economic growth and development by highlighting genetic diversity as a factor that can affect the quality of human capital and the pace of economic progress. In conclusion, our study bridges evolutionary biology and economic theory, highlighting human biological variation as an important factor in explaining disparities in income, education, and wealth.

However, several limitations of our analysis must be acknowledged. First, the UK Biobank sample, while large and highly detailed, is not fully representative of the general population. Participants are volunteers who tend to be healthier and more educated than average, so there may be a healthy volunteer selection bias. This limits the generalizability of our estimates, although it is less likely to invalidate the core finding of a heterosis effect (which is identified from variation within this sample). Second, our measure of genetic diversity—genome-wide heterozygosity—captures broad genetic variation but does not pinpoint which specific genetic loci or regions drive the effects. Notably, heterozygosity in certain gene clusters (for example, those related to immunity or brain development) may have more direct impacts on specific outcomes, an aspect not captured by our aggregate measure. Third, the empirical models we employ (including linear probability specifications with high-dimensional fixed effects for some binary outcomes) could introduce bias or imprecision in those

estimates. Finally, our analysis is essentially cross-sectional, measuring genetic diversity and current outcomes at one point in time. This means we cannot definitively establish causal pathways – for example, we cannot rule out all forms of reverse causality or unobserved environmental factors correlated with genetic diversity. These limitations suggest appropriate caution in interpreting the results and highlight that further empirical research is needed to deepen our understanding of the heterosis effect in human populations.

There are several promising directions for future research. One important step is to replicate and extend these findings in more diverse and representative populations. Examining datasets from other countries or ethnic groups would help assess whether the socioeconomic heterosis effects observed in the UK Biobank generalize more broadly and would strengthen the external validity of our conclusions. Longitudinal studies or intergenerational panels could also shed light on causality and dynamics – for instance, tracking how genetic diversity influences life outcomes over the life course, or whether heterozygosity in parents affects the human capital development of their children. Additionally, more granular genetic analyses are warranted to identify which regions of the genome or which biological pathways underlie the heterosis advantage. It would be illuminating to determine if, say, heterozygosity in immune-function genes improve health (and thereby productivity), or if diversity in neurological genes enhances cognitive flexibility. Investigating gene–environment interactions will likewise be crucial: the payoff to genetic diversity might depend on context, such as the quality of schools, healthcare, or labor market conditions. By integrating approaches from genetics, economics, and other social sciences, future research can better illuminate the mechanisms through which genetic diversity shapes human capital formation and economic outcomes.

In conclusion, our study provides robust evidence that genetic diversity has a positive role in shaping individual human capital and economic success. This insight carries both theoretical and empirical importance. Theoretically, it suggests that economists can enrich human capital and growth models by accounting for biological diversity as a factor that contributes to productivity and innovation. Empirically, it demonstrates through a large-scale analysis that genetic heterogeneity—an aspect of human biology often absent from economic studies—significantly correlates with better life outcomes. Recognizing the heterosis effect in humans opens up new interdisciplinary avenues of inquiry and helps bridge the gap between evolutionary biology and economics. By demonstrating the benefits of genetic diversity for individual achievement and wealth accumulation, our findings introduce a novel perspective on the drivers of economic prosperity in modern societies.

References:

[1] Arnocky, S., Hodges-Simeon, C., Davis, A. C., Desmarais, R., Greenshields, A., Liwski, R., ... & Puts, D. (2021). Heterozygosity of the major histocompatibility complex predicts later self-reported pubertal maturation in men. *Scientific Reports*, *11*(1), 19862.

[2] Ashraf, Q., & Galor, O. (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, *103*(1), 1-46.

[3] Aidoo, M., Terlouw, D. J., Kolczak, M. S., McElroy, P. D., ter Kuile, F. O., Kariuki, S., ... & Udhayakumar, V. (2002). Protective effects of the sickle cell gene against malaria morbidity and mortality. *The Lancet*, 359(9314), 1311-1312.

[4] Baselmans, B. M., Jansen, R., Ip, H. F., van Dongen, J., Abdellaoui, A., van de Weijer, M. P., ... & Bartels, M. (2019). Multivariate genome-wide analyses of the well-being spectrum. *Nature genetics*, 51(3), 445-451.

[5] Barth, D., Papageorge, N. W., & Thom, K. (2020). Genetic endowments and wealth inequality. *Journal of Political Economy*, *128*(4), 1474-1522.

[6] Beauchamp, J.P. (2016). Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences*, 113(28), pp.7774-7779.

[7] Biroli, P., Galama, T., von Hinke, S., Van Kippersluis, H., Rietveld, C. A., & Thom, K. (2025). The economics and econometrics of gene–environment interplay. *Review of Economic Studies*, rdaf034.

[8] Bittles, A. H. (2002). Endogamy, consanguinity and community genetics. *Journal of Genetics*, 81(3), 91-98.

[9] Borgerhoff Mulder, M., & Beheim, B. A. (2011). Understanding the nature of wealth and its effects on human fitness. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *366*(1563), 344-356.

[10]Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., ... & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203-209.

[11] Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-015.

[12] Chen, Z. J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. Nature Reviews Genetics, 14(7), 471-482.

[13] Comings, D. E., & MacMurray, J. P. (2000). Molecular heterosis: a review. Molecular genetics and metabolism, 71(1), 19-31.

[14] Gong, S., Li, Q., Su, S., & Zhang, J. (2024). Genes and sales. *Management Science*, *70*(6), 3902-3922.

[15] Guimaraes, P., & Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *The Stata Journal*, *10*(4), 628-649.

[16] Houmark, M. A., Ronda, V., & 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.

[17] Janicke, T., & Fromonteil, S. (2021). Sexual selection and sexual size dimorphism in animals. *Biology Letters*, *17*(9), 20210251.
[18] Lewis, M. B. (2010). Why are mixed-race people perceived as more attractive?. *Perception*, 39(1), 136-138.
[19]Lind, J. T., & Mehlum, H. (2010). With or without U? The appropriate test for a U-shaped relationship. *Oxford Bulletin of Economics and Statistics*, 72(1), 109-118.
[20] Linnér, R., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., ... & Beauchamp, J. P. (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.
[21] Linnér, R., Mallard, T. T., Barr, P. B., Sanchez-Roige, S., Madole, J. W., Driver, M. N., ... & Dick, D. M. (2021). Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nature neuroscience*, *24*(10), 1367-1376.
[22] Lippman, Z. B., & Zamir, D. (2007). Heterosis: revisiting the magic. *Trends in Genetics*, 23(2), 60-66.
[23] McQuillan, R., Eklund, N., Pirastu, N., Kuningas, M., McEvoy, B. P., Esko, T., ... & ROHgen Consortium. (2012). Evidence of inbreeding depression on human height. *PLoS genetics*, *8*(7), e1002655.
[24] Michod, R. E. (2000). *Darwinian dynamics: evolutionary transitions in fitness and individuality*. Princeton University Press.
[25] Nettle, D., & Pollet, T. V. (2008). Natural selection on male wealth in humans. *The American Naturalist*, *172*(5), 658-666.
[26] Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., ... & Esko, T. (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.
[27] Papageorge, N. W., & Thom, K. (2020). Genes, education, and labor market outcomes: evidence from the health and retirement study. *Journal of the European Economic Association*, *18*(3), 1351-1399.
[28] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559-575.
[29] Rustichini, A., Iacono, W. G., Lee, J. J., & McGue, M. (2023). Educational attainment and intergenerational mobility: a polygenic score analysis. *Journal of Political Economy*, *131*(10), 2724-2779.
[30] Sequeira, T. N., Santos, M., & Ferreira-Lopes, A. (2019). Human capital and genetic diversity. *Eurasian Economic Review*, *9*, 311-330.
[31] Shin, S. H., Lillard, D. R., & Bhattacharya, J. (2020). Understanding the correlation between Alzheimer's disease polygenic risk, wealth, and the composition of wealth holdings. *Biodemography and Social Biology*, *65*(4), 323-350.
[32] Smith, D. J., Escott-Price, V., Davies, G., Bailey, M. E., Colodro-Conde, L., Ward, J., ... & O'Donovan, M. C. (2016). Genome-wide analysis of over 106

000 individuals identifies 9 neuroticism-associated loci. *Molecular Psychiatry*, *21*(6), 749-757.

[33] Song, Z., Li, W. D., Jin, X., Ying, J., Zhang, X., Song, Y., ... & Fan, Q. (2022). Genetics, leadership position, and well-being: An investigation with a large-scale GWAS. *Proceedings of the National Academy of Sciences*, *119*(12), e2114271119.

[34] Syvänen, A. C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, *2*(12), 930-942.

[35] Wang, J., Chen, Q., Chen, G., Li, Y., Kong, G., & Zhu, C. (2020). What is creating the height premium? New evidence from a Mendelian randomization analysis in China. *Plos one*, *15*(4), e0230555.

[36] Yengo, L., Yang, J., Keller, M. C., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2021). Genomic partitioning of inbreeding depression in humans. *The American Journal of Human Genetics*, *108*(8), 1488-1501.

[37] Zhu, C., Zhang, X., Zhao, Q., & Chen, Q. (2018). Hybrid marriages and phenotypic heterosis in offspring: evidence from China. *Economics and Human Biology*, 29, 102-114.