Genetically Informed Estimation of Marginal Returns to Education

Tomeu López-Nieto Veitch * Pietro Biroli † Elisabetta De Cao † Rodrigo Pinto ‡

November 12, 2026§

Abstract

Does higher education amplify or compensate for inequalities associated with genetic predisposition to education? Because individuals with different genetic predispositions also select differently into university, comparisons across genetic groups can conflate differences in the gains from university with differences in who self-selects into it. To separate these returns and selection components, we extend the marginal treatment effect (MTE) framework by aligning, via a mapping function, group-specific returns at a common level of unobserved resistance to attending university, even when the distribution of resistance differs across groups. We apply the framework to the post-war expansion of university access in the United Kingdom, combining UK Biobank data with a newly constructed geocoded dataset of higher-education institutions. We use the reduction in distance to the nearest university between birth and adolescence as an instrument for obtaining a university degree and measure genetic predisposition using the Educational Attainment Polygenic Index (EA PGI). We estimate that returns to a university degree are approximately 0.36 log points higher among individuals in the lowest EA PGI tertile than among those in the highest. Our decomposition suggests that this gap is primarily driven by differences in returns at comparable levels of resistance, with little contribution from differential sorting into university. These findings highlight the potential for expanding access to university to compensate for earnings inequalities associated with genetic predisposition to education.

Keywords: Returns to education, Polygenic Index, Marginal Treatment Effects, UK Biobank
JEL classification: I26, C21, C26, J24

*Department of Statistical Sciences, University of Bologna, Italy. Corresponding author: tomeu.lopeznieto@unibo.it
†Department of Economics, University of Bologna, Italy
‡Department of Economics, University of South Florida, Tampa, US
§This research has been conducted using the UK Biobank Resource under Application Number 74002. This project is funded by grants from (i) the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement (ESSGN 101073237, López-Nieto-Veitch), (ii) the ERC under grant agreement No. 101170376 (HARSH, PI: De Cao), and (iii) the Italian Ministry of Universities and Research (MUR) through PRIN 2022 – PNRR (CUP: J53D23015110001, PI: De Cao). The views expressed in this publication are those of the authors and do not necessarily reflect those of the MUR, European Union, or ESSGN. Neither MUR, the European Union, nor the granting authority or ESSGN can be held responsible for them.

1 Introduction

Education has long been viewed as "the great equalizer of the conditions of men" (Mann, 1848). By expanding access to schooling, societies may allow individuals from less advantaged backgrounds to acquire skills, enter higher-paying occupations, and weaken the link between initial conditions and adult outcomes. Yet this equalising role is not automatic. The distributional effect of education depends on two margins: who is induced to enrol, and how much those who enrol gain. If expanded access reaches disadvantaged individuals with high expected gains, education can compress inequality. If, instead, the gains are larger for individuals who are already advantaged, education may amplify pre-existing differences.

In this paper we ask whether expanded access to university can mitigate earnings inequality associated with genetic predisposition to education: does higher education compensate for or reinforce initial heterogeneity? To answer this question, we examine three interconnected dimensions: (i) whether individuals with different genetic predispositions select differently into university; (ii) whether their returns to a university degree differ at comparable margins; and (iii) how much of the difference in average returns across groups reflects differences in gains rather than educational sorting. To model these margins jointly, we extend the Marginal Treatment Effect (MTE) framework (Heckman and Vytlacil, 2005) to allow observed groups to differ in their distribution of unobserved resistance to attending university—the unobserved factors that make individuals less likely to enrol. We introduce a monotone mapping, which we call the "warp," that aligns group-specific resistance ranks on a common scale. This allows us to compare returns at comparable levels of unobserved resistance and to separate differences in the payoff to a university degree from differences in who selects into university.

Empirically, we estimate these components by combining individual-level genetic and socioeconomic data from 157,109 individuals born between 1945 and 1969 in the UK Biobank with a newly constructed geocoded dataset of higher-education institutions in the United Kingdom (UK). Genetic data are useful for this question because they provide a predetermined source of individual heterogeneity that is objectively measured in the same way across space, time, and schooling institutions. We use the Educational Attainment Polygenic Index (EA PGI), a standardised summary measure of genetic variants associated with years of schooling and a broad range of developmental and socioeconomic outcomes (Lee et al., 2018; Okbay et al., 2022; Barth, Papageorge, and Thom, 2020; Belsky et al., 2016), as a predetermined measure of genetic predisposition to education. Our identification strategy exploits the post-war expansion of universities in the UK, leveraging the reduction in distance to the nearest university between birth and adolescence as an instrument for obtaining a university degree.

We estimate an average wage return to a university degree of 0.491 log points, with substantial heterogeneity across individuals.[1] First, individuals with lower unobserved resistance to university obtain larger gains from attendance, consistent with selection on gains (Heckman, Urzua, and Vytlacil, 2006). Second, returns are substantially larger for individuals with low EA PGI than for individuals with

[1] For descriptive context, Blundell, Green, and Jin (2016) report a raw median hourly-wage gap between graduates and school-leavers of 45 percent at age 30 and 58 percent at age 35 for the 1965–69 birth cohort.

high EA PGI. Specifically, over the available common support, the estimated average treatment effect is about 0.687 log points for the low EA PGI group, 0.543 for the middle EA PGI group, and 0.329 for the high EA PGI group. Although we find large differences in selection patterns across EA PGI groups, our decomposition indicates that these differences make a limited contribution to the gap in average returns: the gap is driven primarily by differences in the payoff to a university degree at comparable resistance margins. Overall, the evidence is consistent with university playing a compensatory role in earnings inequalities associated with genetic predisposition to education.

Because individuals select into university partly on the basis of their gains, and educational sorting differs across EA PGI groups, estimating gene–environment interactions ($G \times E$)—how the causal payoff to education varies with genetic predisposition—is not straightforward. In particular, as Hollenbach, Schmitz, and Westphal (2026) show, when complier populations differ across groups, standard IV interaction estimates can mix differences in causal gains with differences in the types of individuals whose educational choices respond to the instrument.[2] To address this issue, they estimate MTE curves separately within genetic groups and obtain interaction effects by comparing marginal returns at the same within-group rank, or percentile, of resistance. In doing so, these quantile-to-quantile comparisons address the problem arising from averaging returns over different complier populations.

However, we identify a further comparability problem. When MTE curves are estimated separately by group, individuals are ranked relative to their own group's resistance distribution. If these distributions differ, the same within-group percentile need not correspond to the same underlying level of resistance. In our setting, genetic groups differ not only in their returns to a university degree but also in their distribution of latent resistance to attending. The median individual in a group with generally low resistance may, for example, face much lower latent costs of attending university than the median individual in a group with generally high resistance. Comparing the two medians therefore compares returns at different points along the underlying resistance distribution. The resulting gap can reflect not only differences in the payoff to a university degree across genetic groups, but also differences in where the individuals being compared lie along that distribution.

Our methodological contribution is to address this comparability problem directly. We extend the MTE framework to link resistance ranks across groups defined by a categorical moderator while allowing their latent resistance distributions to differ.[3] We introduce a monotone mapping function, the warp, that links group-specific resistance ranks to a common pooled resistance scale. This allows comparisons of marginal returns at a common underlying level of resistance and yields a representation of the pooled MTE as an appropriately weighted average of the aligned group-specific MTEs. Our approach builds on work using MTEs to recover economically relevant treatment-effect parameters (Mogstad, Santos, and Torgovitsky, 2018) and relates to research showing how differences in selection can reconcile treatment-effect estimates

[2] Under the usual LATE assumptions, group-specific IV estimates identify average gains for the compliers within each group. When individuals select on gains, differences in the composition of these complier populations can translate into differences in their average returns. Comparing the estimates therefore need not isolate heterogeneity in the causal payoff to education across comparable selection margins.

[3] The framework applies to treatment-effect heterogeneity across observed groups more generally. For example, it can be used to compare returns to a university degree between women and men at a common resistance margin and to decompose differences in average returns into returns and selection-related components.

across settings (Kowalski, 2023).[4]

Building on this alignment, we ask how much of the difference in average returns across groups reflects differences in the gains from university and how much reflects differences in educational sorting. We decompose between-group differences in treatment-effect parameters into (i) a returns component and (ii) selection-related components. The first captures differences between the aligned MTE curves, reflecting different causal payoffs to a university degree at a common resistance margin. The latter capture differences in the distribution of resistance and in who selects into university, which determine how marginal gains are weighted when computing average returns for a given population.[5] In our application, the returns component captures gene–environment interaction ($G \times E$), whereas the selection-related channels capture the fact that genetic predisposition shapes the educational environments individuals select into, connecting naturally to active gene–environment correlation (rGE).[6]

Our application contributes to the growing economics literature using genetic data to study human capital formation and inequality (Benjamin et al., 2026; Biroli et al., 2026). This literature provides new ways to examine how initial endowments interact with the environments and institutions individuals encounter, a question central to debates about the economic consequences of the "genetic lottery" (Harden, 2021). For example, Papageorge and Thom (2020) show that the association between the EA PGI and university graduation is stronger among individuals from more advantaged childhood backgrounds. Their findings highlight how family circumstances shape the relationship between genetic predisposition and educational attainment. Turning from educational attainment to its causal consequences, the closest empirical study is Barcellos, Carvalho, and Turley (2021), who examine whether an additional year of compulsory schooling induced by the UK Raising of the School Leaving Age reform changes the relationship between the EA PGI, early-life disadvantage, and later-life socioeconomic outcomes. We extend this evidence to university, a central margin of human-capital investment that shapes access to skilled occupations and lifetime earnings opportunities.

By studying genetic heterogeneity in the economic returns to a university degree, we also contribute to the longstanding literature on how educational investments shape earnings and inequality.[7] Carneiro, Heckman, and Vytlacil (2011) show how selection on gains makes the returns to marginal expansions in university attendance depend on which individuals are induced to enrol.[8] Nybom (2017) finds that lifetime returns to a university degree vary substantially with cognitive and noncognitive skills, suggesting complementarities between education and these skills, but relatively little with parental earnings. Evidence from university admission thresholds also shows that university can generate substantial earnings

[4] For broader discussions of MTE methods, identification, and policy evaluation, see Cornelissen et al. (2016); Mogstad and Torgovitsky (2018, 2024); see also Brinch, Mogstad, and Wiswall (2017) for related identification results.

[5] The decomposition applies to the average treatment effect and to average effects for treated and untreated individuals, with the relevant aggregation weights depending on the parameter of interest. We use a Shapley decomposition to allocate the contributions without privileging a particular ordering (Shorrocks, 2013).

[6] Active gene–environment correlation refers to individuals selecting environments in ways related to their genetic predispositions.

[7] Early contributions include Renshaw (1960); Becker (1964); Mincer (1974); see Psacharopoulos (1994); Card (1999) for influential reviews of the evidence on returns to schooling.

[8] At the distributional level, Carneiro and Lee (2009) show how changes in university enrolment can affect wage inequality by changing the composition of educational groups, even holding skill prices fixed.

gains for academically marginal students (Zimmerman, 2014; Mountjoy, 2026). This evidence shifts attention from the average payoff to a university degree towards the gains of individuals whose attendance changes when access expands.[9]

Our empirical design connects this question to a growing literature exploiting expansions in local higher-education supply to study the consequences of university access. Westphal, Kamhöfer, and Schmitz (2022) estimate marginal wage returns to university education using university expansion in West Germany while accounting for selection into employment, and Braccioli et al. (2024) examine how expansion in Italy affects educational choices and subsequent labour-market outcomes.[10] In the UK, Blundell, Green, and Jin (2022) document a broadly stable graduate wage premium during the substantial expansion of graduate supply between 1993 and 2015, highlighting how firms' organisational responses can sustain the premium as participation grows. Moving from the aggregate premium to differences across individuals, Ichino, Rustichini, and Zanella (2026) examine how UK university expansion changed university participation and earnings premiums across cognitive-ability and socioeconomic groups. They find that expansion increasingly drew students with lower cognitive ability from advantaged backgrounds into university, and identify alternative policies that could have better reached high-ability students from disadvantaged backgrounds. We complement this evidence by examining how the gains from expanded access to university vary with genetic predisposition to educational attainment.[11]

Hence, whether university expansion strengthens education's equalising role depends on both whose access improves and how much those individuals gain from attending. Our findings highlight the potential for expanded access to education to weaken the link between initial endowments and adult economic outcomes. The rest of the paper is structured as follows. Section 2 presents the theoretical framework. Section 3 describes the data and empirical setting. Section 4 reports the main empirical results and decomposition analysis. Section 5 concludes.

2 Roy Model with a Categorical Moderator

2.1 Setup

We develop a modified Roy model (Roy, 1951) building on the marginal treatment effects (MTE) framework of Heckman and Vytlacil (2005). The key feature is an observed categorical variable $G \in \mathcal{G} = \{g_1, \ldots, g_K\}$ that shifts the distribution of latent heterogeneity relevant for both treatment selection and potential outcomes. Let $E \in \{0, 1\}$ denote a binary treatment, Y an outcome, and Z a continuous

[9] The relevant educational alternative also matters. Mountjoy (2022) show that community-college access benefits students who would otherwise not enter higher education but can reduce outcomes for those diverted from immediate four-year entry. At the field-of-study margin, Osikominu, Pfeifer, and Ruberg (2026) use MTEs to document selection on gains into STEM and show how the returns to policies expanding STEM enrolment depend on the individuals affected.

[10] Related work examines a wider range of consequences of higher-education access: Currie and Moretti (2003) study maternal education and infant health using US college openings; Kamhöfer, Schmitz, and Westphal (2019) examine heterogeneous non-monetary returns to higher education in Germany; Suhonen and Karhunen (2019) use the geographical expansion of Finnish universities to study the intergenerational effects of parental education; and Berlingieri, Gathmann, and Quinckhardt (2022) investigate university openings and local economic development.

[11] The EA PGI, measured cognitive ability, and socioeconomic background capture distinct dimensions of heterogeneity.

instrument shifting the propensity to receive treatment.

In our application, E corresponds to university degree attainment, Y to (log) wages, and G to categories of genetic predisposition to educational attainment. We adopt the potential outcomes notation and let Y_1 and Y_0 denote the potential outcomes under treatment and no treatment, respectively, so that the observed outcome is:

$$Y = E\,Y_1 + (1 - E)\,Y_0$$

Let $\boldsymbol{V} \in \mathcal{V} \subseteq \mathbb{R}^d$ denote a random vector (of arbitrary dimensions, d) of latent confounders, that affect both selection and outcomes. In the returns to education, $\boldsymbol{V}$ may capture factors such as ability, preferences and other unobserved determinants of schooling choice and later outcomes, with G shifting the distribution of this heterogeneity. Then, the model is represented by the following structural equations:

$$\textbf{Confounders}: \; \boldsymbol{V} = f_V(G, \varepsilon_V) \tag{1}$$

$$\textbf{Selection equation}: \; E = f_E(Z, \boldsymbol{V}) \tag{2}$$

$$\textbf{Outcome equation}: \; Y = f_Y(E, \boldsymbol{V}, \varepsilon_Y) \tag{3}$$

Here, ε_V and ε_Y denote idiosyncratic shocks and potential outcomes are defined by $Y_e = f_Y(e, \boldsymbol{V}, \varepsilon_Y)$ for any treatment state, $e \in \{0, 1\}$. Figure 1 summarises the assumed causal structure of the model. We do not include G directly in the choice or outcome equations. Instead, to capture all possible channels through which G may affect both treatment and outcome, we let G operate through the vector of confounding unobservables $\boldsymbol{V}$[12]. For notational simplicity, we omit control variables $\boldsymbol{X}$, although all results extend immediately to the case in which conditioning on $\boldsymbol{X}$ is made explicit throughout.

In this modified Roy model, the following standard instrumental-variable assumptions must hold:

(A1) Relevance: Z affects E

(A2) Exclusion: Z affects Y only through E

(A3) Exogeneity: $Z \perp (G, \varepsilon_Y, \varepsilon_V)$ and $G \perp (\varepsilon_Y, \varepsilon_V)$

(A4) Monotonicity: Z shifts E in a common direction

(A5) $Z, \boldsymbol{V}$ are absolutely continuous

Since $\boldsymbol{V} \in \mathbb{R}^d$ is a random vector, for any Borel-measurable function $\psi : \mathbb{R}^d \to \mathbb{R}$, we have that $S \equiv \psi(\boldsymbol{V})$ is a real-valued random variable, which describes a (scalar) latent index based on unobservables. Therefore, the cumulative distribution function (CDF) of S exists:

[12] This formulation is without loss of generality, since V is of arbitrary dimension.

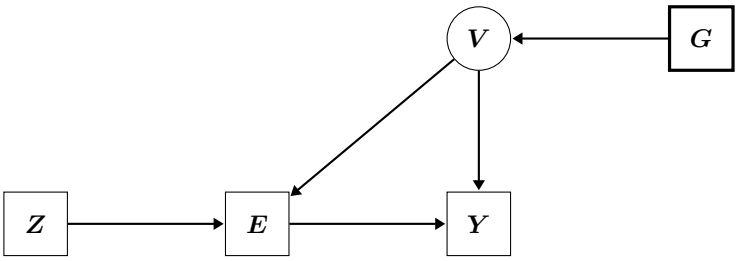


Figure 1: Directed acyclic graph (DAG) of the general IV model with group G. The instrument, Z, affects treatment, E, and affects outcome, Y, only through E. Unobserved heterogeneity $\boldsymbol{V}$ affects both E and Y and G shifts the distribution of $\boldsymbol{V}$

$$F_S(s) = \Pr(S \leq s)$$

Likewise, for each group $g \in \mathcal{G}$ we can define the group-specific CDF as,

$$F_S^g(s) = \Pr(S \leq s \mid G = g)$$

Assume further that S is continuously distributed in the pooled population and within each group $g \in \mathcal{G}$. Then, by the probability integral transform, we can rank individuals based on their unobservables as:

$$U \equiv F_S(S) \sim \text{Unif}(0,1) \tag{4}$$

$$U^g \equiv F_S^g(S) \,\big|\, G = g \sim \text{Unif}(0,1) \tag{5}$$

Following Heckman and Vytlacil (2005), we assume that, for any Borel-measurable $\zeta : \mathbb{R} \to \mathbb{R}$ mapping function, selection into treatment E can be written as a latent threshold-crossing equation:

$$E = \mathbf{1}\{\zeta(Z) \geq S\}$$

Under this threshold-crossing representation, S admits an interpretation as latent resistance (or distaste) to treatment. Individuals with higher values of S require a larger instrument-induced incentive $\zeta(Z)$ to select into treatment. Accordingly, U can be interpreted as the quantile of unobserved resistance in the pooled distribution, and U^g as the corresponding quantile in the group-specific distribution.

Let $P(Z) \equiv \Pr(E = 1 \mid Z)$ denote the pooled propensity score and let $P^g(Z) \equiv \Pr(E = 1 \mid Z, G = g)$

denote the group-specific propensity score. The treatment rules can be expressed as[13]:

$$E = \mathbf{1}\{P(Z) \geq U\} \tag{6}$$
$$E = \mathbf{1}\{P^g(Z) \geq U^g\}$$

where $P(Z) \perp U$ and $P^g(Z) \perp U^g$ conditional on $G = g$.

By (6), individuals with $U = P(Z)$ are those at the margin of indifference between treatment and non-treatment in the pooled population, and analogously individuals with $U^g = P^g(Z)$ are at the margin within group g. Therefore, the Marginal Treatment Effect (MTE) is defined as the average gain from treatment for individuals at a given level of unobserved resistance:

$$\text{MTE}(u) = \mathbb{E}(Y_1 - Y_0 \,|\, U = u)$$

Analogously, for any $u^g \in U^g$, the group-specific MTE is defined as:

$$\text{MTE}^g(u^g) = \mathbb{E}(Y_1 - Y_0 \,|\, U^g = u^g, G = g)$$

2.2 The Problem: Comparability of Group-Specific Resistance Ranks

A crucial implication of the previous subsection is that the pooled resistance rank U and the group-specific resistance rank U^g are distinct random variables. They are both constructed from the same latent index S, but through different CDFs: $U = F_S(S)$ and $U^g = F_S^g(S)$. Thus, U ranks individuals in the pooled distribution of latent resistance, whereas U^g ranks them within group g. As a result, the same latent resistance level need not map into the same numerical rank in the pooled and group-specific distributions (i.e., $F_S(S) \neq F_S^g(S)$), which can lead to misleading comparisons across groups.

We are interested in comparing group-specific marginal treatment effects, MTE^g, in order to assess heterogeneity in treatment effects across groups. In our application, for instance, we want to compare the returns to a university degree between individuals with a high genetic predisposition to educational attainment and individuals with a low genetic predisposition. As an illustration, suppose that individuals with a high genetic predisposition ($G = H$) are, ceteris paribus, more likely to select into university than individuals with a low predisposition ($G = L$). In this case, individuals in group H are, on average, less resistant to university. Figure 2 illustrates this idea. At a given latent resistance level s_0, a larger proportion of individuals in the high-predisposition group will have resistance below s_0 than individuals in the low-predisposition group, so that $F_S^H(s_0) > F_S(s_0) > F_S^L(s_0)$.

Hence, the same latent resistance level corresponds to different percentile ranks in the H and L groups. Conversely, fixing a common numerical rank across groups need not correspond to fixing the same underlying resistance level. This matters directly for the interpretation of group-specific MTEs. Suppose, for example, that we compare $\text{MTE}^H(0.5)$ and $\text{MTE}^L(0.5)$ to ask whether the returns to a university

[13] $\Pr(E = 1|Z) = \Pr(S \leq \zeta(Z)) = \Pr(F_S(S)) \leq F_S(\zeta(Z))) = \Pr(U \leq F_S(\zeta(Z))) = F_S(\zeta(Z))$. The first equality holds due to $Z \perp S$, the second one uses the fact that $F_S(\cdot)$ is strictly increasing, and the last one holds because U is uniformly distributed.

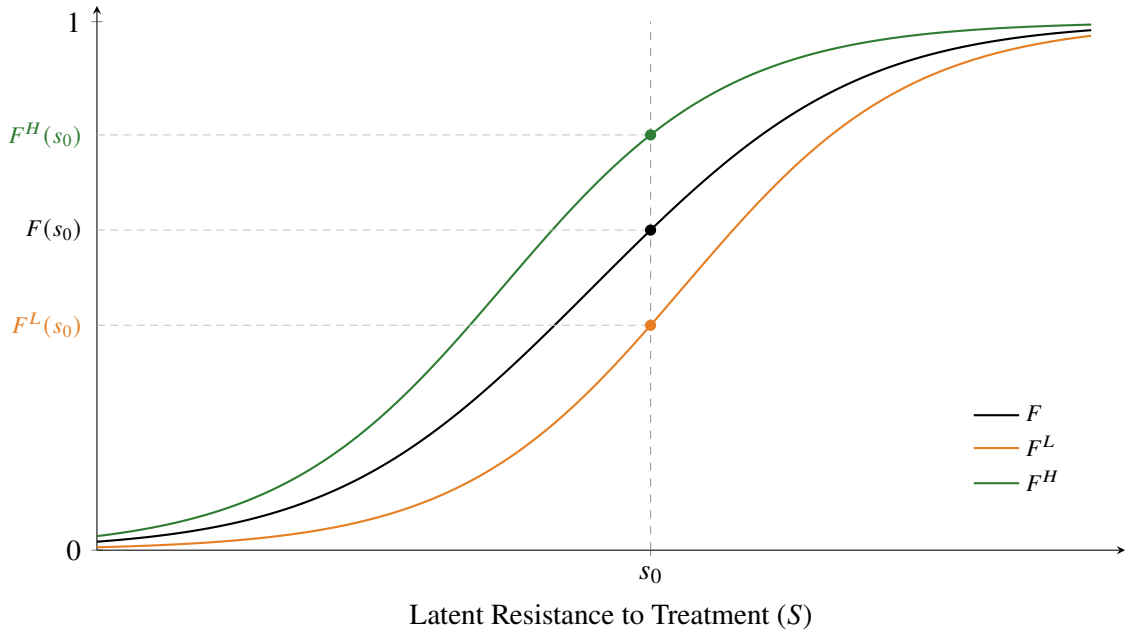


Figure 2: Illustration of rank misalignment across groups. The figure plots the pooled and group-specific cumulative distribution functions of latent resistance S for the pooled population (F), the high-predisposition group (F^H) and the low-predisposition group (F^L). At the common resistance level s_0, the corresponding percentile ranks differ across distributions, with $F_S^H(s_0) > F_S(s_0) > F_S^L(s_0)$.

degree are higher for individuals with high rather than low genetic predisposition at the median of latent resistance ($u = 0.5$). Such a comparison is not straightforward. It compares individuals who occupy the same quantile within their respective group distributions, but who may face very different underlying latent costs of obtaining a university degree.

More generally, the issue is not limited to point-wise MTE comparisons at a common pooled rank. Differences arising from the comparison of treatment effects across groups may reflect not only heterogeneity in returns, but also differences in selection patterns. That is, differences in who is located at each rank within each group[14]. The resulting difference may therefore combine two distinct sources of variation: genuine heterogeneity in treatment gains and differences in where each group lies in the latent resistance distribution.

[14]This relates to, but differs from, the concern emphasised by Hollenbach, Schmitz, and Westphal (2026), who show that 2SLS interaction estimates may be misleading when groups differ in their complier populations, and propose comparing MTEs to maintain a fixed underlying population. However, even once one moves to group-specific MTEs, comparability is still not automatic if each group-specific choice equation induces its own group-specific resistance rank. In that case, equal subgroup quantiles are only quantile-to-quantile comparisons, not necessarily comparisons at a common latent resistance level.

2.3 The Solution: Mapping from Group-Specific to Pooled Resistance

To compare treatment effects across groups at a common level of latent resistance, we need a mapping bridging the pooled rank U and the group-specific rank U^g[15]. The key object is a *warp* function that takes a pooled resistance rank u and returns the percentile rank that the same underlying resistance level occupies within group g.

Definition 1 (**The Warp Function**). *For each group $g \in \mathcal{G}$, the warp function $u^g : [0,1] \to [0,1]$ is defined as the composite mapping:*

$$u^g(u) \equiv F_S^g(F_S^{-1}(u)) \tag{7}$$

Thus, $u^g(\cdot)$ is a quantile-to-quantile map. It starts from a pooled rank u, recovers the associated resistance level $s = F_S^{-1}(u)$, and then evaluates where that same resistance level lies in the distribution of group g. The warp function therefore puts all groups on a common resistance scale. Equivalently, $u^g(u)$ gives the within-group percentile rank associated with the same underlying resistance level as pooled rank u.

Figure 3 illustrates this mapping. A warp above the 45-degree line implies that, at a given pooled rank, individuals in group g occupy a higher percentile within their own group distribution ($u^g(u) > u$). This corresponds to a group that is less resistant, on average, at that common resistance margin. By contrast, a warp below the 45-degree line indicates that the same pooled rank corresponds to a lower within-group percentile ($u^g(u) < u$), consistent with greater resistance. In the figure, the pooled 60th percentile of resistance, $u = 0.6$, maps to the 80th percentile ($u^H(0.6) = 0.8$) in the high genetic predisposition group and to the 20th percentile in the low predisposition group ($u^L(0.6) = 0.2$). This shows that the same common margin can correspond to very different positions in the group-specific resistance distributions.

The following proposition lists the essential properties of this mapping:

Proposition 1 (**Properties of the Warp**). *Let F and $F^g; g \in \mathcal{G}$ be continuous and strictly increasing on a common support, with densities $f > 0$ and $f^g > 0$ on the interior. Then:*

(i) Monotonicity and derivative. *u^g is strictly increasing and*

$$\omega_g(u) \;=\; (u^g)'(u) \;=\; \frac{f^g(F^{-1}(u))}{f(F^{-1}(u))} \;>\; 0.$$

(ii) Endpoints. *$u^g(0) = 0$ and $u^g(1) = 1$ under common support and no boundary atoms.*

(iii) Mixture (Barycentric) Identity. *$\sum_g \pi_g\, u^g(u) \;=\; u$, for all $u \in [0,1]$, where $\pi_g \equiv \Pr(G = g)$ and $\sum_g \pi_g = 1$.*

(iv) Calibration identity. *$u^g\big(P(Z)\big) \;=\; P^g(Z)$ a.s. in Z, under the threshold-crossing representation (6).*

[15]Note that since the pooled distribution is a mixture of the group-specific distributions, $F_S(s) = \sum_{g\in\mathcal{G}} \Pr(G = g)\, F_S^g(s)$, the mapping between U and U^g is generally nonlinear whenever $F_S^g \neq F_S$.

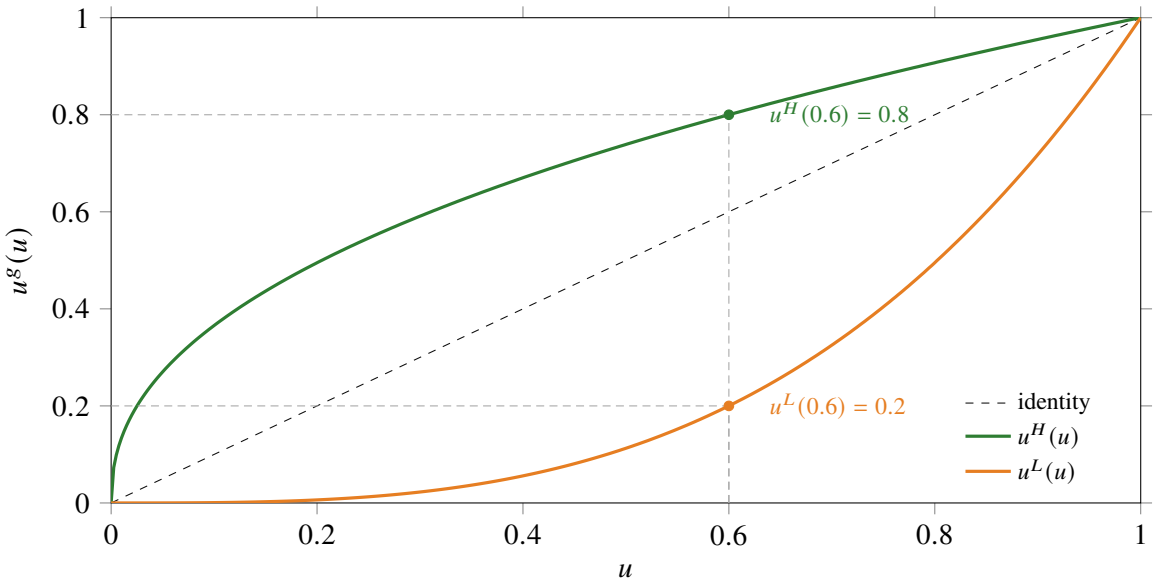


Figure 3: Illustration of the warp function. The figure shows two illustrative warp mappings from a pooled rank u into group-specific ranks $u^g(u)$. The 45-degree line corresponds to no rank misalignment. A curve above the identity implies that, at a given pooled rank, individuals occupy a higher percentile rank within that group. A curve below the identity implies a lower within-group rank. In the example, the same pooled rank $u = 0.6$ maps into $u^H(0.6) = 0.8$ for Group H and into $u^L(0.6) = 0.2$ for Group L, illustrating how the same common margin can correspond to very different positions in the group-specific resistance distributions.

(v) Posterior composition. *The posterior group probability at pooled rank u satisfies:*

$$\Pr(G = g \mid U = u) \;=\; \pi_g\,(u^g)'(u) = \pi_g \omega_g(u), \text{ and therefore } \sum_g \pi_g\,\omega_g(u) \;=\; 1.$$

Item (*i*) states that the warp $u^g(u)$ is monotone and that its (Radon-Nikodym) derivative, the *tilt* function $\omega_g(u)$ is a local density ratio comparing the group-specific and pooled resistance distributions at the same underlying margin. Item (*ii*) states that, under full support, the warp function has well-behaved endpoints. (i) and (ii) together, formalise that u^g is an order-preserving quantile–to–quantile map. Item (*iii*) shows that the weighted average of the group-specific warp functions reproduces the identity, reflecting the fact that the pooled distribution is a mixture of the group-specific distributions. Item (*iv*) identifies the warp from first-stage propensity objects, linking the warp directly to observables. The group-specific propensity score is equivalent to the warped version of the pooled propensity score. Finally, item (*v*) shows that the tilt function also has a composition interpretation, since $\pi_g\omega_g(u)$ is the posterior probability of belonging to group g at pooled rank u.

2.4 Treatment Effects

Having mapped group-specific resistances (u^g) onto the common pooled resistance scale (u), we can now express group-specific treatment parameters in a form that is directly comparable across groups. As

described in Heckman and Vytlacil (2005), any targeted treatment effect of interest τ (e.g. ATE, ATT, ATU, IV) can be computed as a weighted average of the MTE:

$$\tau = \int_0^1 \text{MTE}(u)\, w_\tau(u) du$$

where w_τ is a non-negative weight associated to the target estimand τ. Analogously, the group-specific treatment effect, τ^g, can be written as a weighted average of the group-specific MTE over its group-specific rank $v \in [0, 1]$:

$$\tau^g = \int_0^1 \text{MTE}^g(v)\, w_\tau^g(v) dv$$

To compare τ^g and $\tau^{g'}$ across groups, however, these group-specific treatment effects must be projected onto the common pooled-rank domain $u \in [0, 1]$. This is done in the following proposition:

Proposition 2 (**Transport of group-specific treatment functionals**). *Every group-specific treatment functional admits the pooled-rank representation:*

$$\begin{aligned} \tau^g &= \int_0^1 MTE^g(v)\, w_\tau^g(v)\, dv = \int_0^1 MTE^g\big(u^g(u)\big)\, w_\tau^g\big(u^g(u)\big)\, \omega_g(u)\, du, \\ \textit{where } u^g(u) &\equiv F_S^g\big(F_S^{-1}(u)\big) \quad \textit{is the } \text{warp } \textit{function}, \\ \textit{and } \omega_g(u) &\equiv \frac{f_S^g(F_S^{-1}(u))}{f_S(F_S^{-1}(u))} \quad \textit{is the } \text{tilt } \textit{weight.} \end{aligned}$$

Proposition 2 shows that once a group-specific treatment functional is projected onto the pooled rank domain, its aggregation depends on three distinct components: the group-specific MTE curve $\text{MTE}^g(\cdot)$, the estimand-specific weight $w_\tau^g(\cdot)$, and the distributional objects $u^g(\cdot)$ and $\omega_g(\cdot)$. The warp re-indexes native group ranks onto the common pooled scale, while the tilt appears as the Jacobian term induced by this change of variables.

Corollary 1 (**Reconstruction of the pooled MTE**). *The pooled MTE admits the representation*

$$\text{MTE}(u) = \sum_{g \in \mathcal{G}} \pi_g\, \text{MTE}^g\big(u^g(u)\big)\, \omega_g(u), \tag{8}$$

where $\pi_g \equiv \Pr(G = g)$.

Corollary 1 makes clear that the pooled MTE is not a simple average of native group-specific MTE curves. Rather, each group-specific MTE must first be evaluated at the warped rank $u^g(u)$ and then weighted by the local composition term $\pi_g \omega_g(u)$.

Corollary 2 (Pooled treatment-effect decomposition). *Any pooled treatment effect parameter satisfies*

$$\tau = \int_0^1 \mathrm{MTE}(u)\, w_\tau(u)\, du = \sum_{g \in \mathcal{G}} \pi_g \int_0^1 \mathrm{MTE}^g\big(u^g(u)\big)\, \omega_g(u)\, w_\tau(u)\, du.$$

In particular, for the average treatment effect,

$$\begin{aligned}\mathrm{ATE} &= \int_0^1 \mathrm{MTE}(u)\, du \\ &= \sum_{g \in \mathcal{G}} \pi_g \int_0^1 \mathrm{MTE}^g\big(u^g(u)\big)\, \omega_g(u)\, du \\ &= \sum_{g \in \mathcal{G}} \pi_g\, \mathrm{ATE}^g.\end{aligned}$$

Corollary 2 shows that pooled treatment-effect parameters can be decomposed into group-specific components expressed on the common pooled resistance scale. For generic treatment parameters, the decomposition depends on the estimand-specific weight $w_\tau(u)$. The ATE is a special case in which this simplifies to a mixture of group-specific ATEs with population weights π_g.

2.5 Decomposition of cross-group differences of treatment effects

We now turn to the decomposition of cross-group differences in treatment effects, $\tau^g - \tau^{g'}$, for any two groups $g, g' \in \mathcal{G}$. As highlighted in Section 2.2, such differences need not reflect pure heterogeneity in causal returns. In general, they may combine differences in marginal treatment effects with differences in how those marginal effects are aggregated across latent resistance ranks. We aim to disentangle the latter. To formalise this decomposition, let

$$A_g(u) \equiv \mathrm{MTE}^g\big(u^g(u)\big)$$

denote the group-specific marginal effect evaluated on the common pooled-rank domain, and let

$$W_g(u) \equiv B_g(u)\, C_g(u)$$

denote the corresponding weighting term, where

$$B_g(u) \equiv w_\tau^g\big(u^g(u)\big) \qquad \text{and} \qquad C_g(u) \equiv \omega_g(u).$$

Here, $B_g(u)$ is the *target weight*, which determines how the estimand of interest aggregates marginal effects across resistance margins, while $C_g(u)$ is the *tilt*, which captures local differences in group composition along the pooled resistance scale. In this notation, the group-specific treatment functional

can be written as

$$\tau^g = \int_0^1 A_g(u)\, B_g(u)\, C_g(u)\, du.$$

The following proposition decomposes $\tau^g - \tau^{g'}$ into three interpretable sources: differences in marginal effects, differences in target weights, and differences in composition.

Proposition 3. *The difference $\tau^g - \tau^{g'}$ admits the following exact decomposition:*

$$\begin{aligned}
\tau^g - \tau^{g'} = &\int_0^1 \left(\frac{W_g + W_{g'}}{2}\right)(A_g - A_{g'})\, du && \text{(MTE Channel)} \\
&+ \int_0^1 \left(\frac{A_g + A_{g'}}{2} \cdot \frac{C_g + C_{g'}}{2}\right)(B_g - B_{g'})\, du && \text{(Target Weight Channel)} \\
&+ \int_0^1 \left(\frac{A_g + A_{g'}}{2} \cdot \frac{B_g + B_{g'}}{2}\right)(C_g - C_{g'})\, du. && \text{(Composition Channel)}
\end{aligned}$$

This decomposition is obtained through two nested two-way Shapley steps. First, we decompose $\tau^g = \int A_g W_g\, du$ into the marginal-effect component A_g and the aggregate weighting component W_g. Second, decomposing $W_g = B_g C_g$ into the target-weight component B_g and the composition component C_g. Each channel isolates the difference in one factor ($A_g - A_{g'}$, $B_g - B_{g'}$, or $C_g - C_{g'}$), while averaging the remaining factors symmetrically across groups.

The first term, associated with differences in $A_g(u)$, is a *returns channel*. It captures how treatment gains differ across groups at a common resistance margin. More generally, it reflects differences in the causal response to treatment holding the resistance margin fixed. Thus, it can also be interpreted as a technology or behavioural-response channel. The second and third terms, associated with differences in $B_g(u)$ and $C_g(u)$, together form a *selection channel*. They capture how differences in treatment-effect parameters arise not only from differences in returns, but also from differences in how marginal gains are aggregated across resistance margins and from differences in who is located at each margin.

These interpretations are especially natural in our application, where groups are defined by genetic predisposition to educational attainment. In this setting, the returns channel can be interpreted as a form of gene–environment interaction (GxE), since it captures heterogeneity in the causal returns to education across genetic groups at a common level of latent resistance. By contrast, the target-weight and composition channels capture differences in educational selection across genetic groups. For treatment parameters whose weights depend only on propensity-score objects, such as the ATE, ATT and ATU, these channels can be interpreted as reflecting (active) gene–environment correlation (rGE), insofar as genetic predisposition is related to selection into education.

Finally, a few special cases are worth noting. For the ATE, $B_g = B_{g'} \equiv 1\,, \forall g \in \mathcal{G}$, so the target-weight channel vanishes and the difference depends only on the returns component, A, and the composition component, C. If groups have no impact on marginal effects, $A_g = A_{g'}$, then all cross-group differences are driven by target-weight differences B and/or composition differences C. Under no composition heterogeneity, $F^g = F$ for all g, we have $u^g = u^{g'}$ and $C_g = C_{g'} = 1$, so the decomposition reduces to

differences in marginal effects and target weights alone.

2.6 Identification and Estimation

Marginal Treatment Effects

We estimate marginal treatment effects using the Local IV (LIV) approach of Heckman and Vytlacil (1999). Under the standard latent-index selection model and the usual IV assumptions, including monotonicity and sufficient support, the MTE at resistance rank u is identified as the derivative of the conditional mean of outcomes with respect to the propensity score:

$$\text{MTE}(x,u) = \left.\frac{\partial\,\mathbb{E}[Y \mid X = x,\ P(Z) = p]}{\partial p}\right|_{p=u} \tag{9}$$

We implement this in three steps. First, we estimate the propensity score $\widehat{P}(Z) = \widehat{\Pr}(E = 1 \mid Z)$ as a Probit[16] and restrict estimation to the trimmed region of common support between treated and untreated individuals. Second, we separately partial out the nonparametric dependence of Y, each covariate in X, and the interaction terms $X'\widehat{P}(Z)$ on the estimated propensity score using local linear smoothers.[17] We then regress the residualised outcome on the residualised covariates and interaction terms to estimate the coefficients of the observable component of the outcome equation. This yields an estimate of the observable contribution to outcomes. We subtract this fitted contribution from Y to obtain the residual outcome, $\tilde{Y}$. Third, we estimate the conditional expectation of this residual outcome, $\mathbb{E}[\tilde{Y} \mid \widehat{P}(Z) = p]$, nonparametrically and estimate its derivative with respect to p by fitting a local quadratic regression with a Gaussian kernel.[18] Then, we construct $\widehat{\text{MTE}}(x,u)$ by combining the estimated derivative with the observable treatment-effect component.[19]

To estimate the group-specific $\text{MTE}^g(x,u)$, we repeat the LIV procedure separately within each genetic group $G = g$. This allows the MTE function to vary flexibly across genetic strata, thereby accommodating rich heterogeneity in both observables and unobservables. Therefore, we first estimate objects $\widehat{P}^g(Z)$ and obtain the equivalent group-specific objects such as $\tilde{Y}^g$. Our warp transformation then maps the resulting group-specific MTEs onto a common resistance scale, u. We do not impose additive separability between G and the unobserved resistance term U, nor parallel-slope restrictions across genetic groups, as is common in bounded-MTE approaches developed for discrete or limited-support instruments (e.g., Hollenbach, Schmitz, and Westphal (2026)).

[16] Note that Z may include some or all of the components in X, but at least one variable not included in the outcome equation. We estimate the propensity as a Probit, like e.g., Cornelissen et al. (2018). In Table A.11 we show that results are robust to the choice of alternative first stage link functions (e.g. Logit).

[17] We implement the local linear smoothers using LOESS with span 0.10.

[18] Local quadratic estimators are generally recommended to fit first-order derivatives (Fan and Gijbels, 1996). We use a Gaussian kernel with bandwidth 0.15, assessed against earlier five-fold cross-validation results.

[19] We estimate the model using our own modified and extended implementation building on the `localIV` R package (Zhou, 2020).

Warp and the Tilt Weights

From Proposition 1 we have that the identification of the warp, $u^g(u)$, relies on the calibration identity $u^g(P(Z)) = P^g(Z)$. Hence, $u^g(u) : [0,1] \to [0,1]$ is the monotone regression function of $P^g(Z)$ on $P(Z)$ and can be identified by the following regression:

$$\hat{u}^g(\cdot) \equiv \arg\min_{m \in \mathcal{M}} \sum_{i \in \mathcal{V}^g} \left(\hat{P}_i(Z)^g - m(\hat{P}_i(Z)) \right)^2 \tag{10}$$

where $\mathcal{M}$ is the class of monotone functions $m : [0,1] \to [0,1]$ and $\hat{P}_i(Z)^g$, $\hat{P}_i(Z)$ are the estimated conditional and unconditional propensity score values, respectively, for observations in $\mathcal{V}_g : \{i : G_i = g\}$. In our application, we estimate $m(\cdot)$ via Isotonic Regression. This approach is particularly well-suited for the warp function as it provides a non-parametric fit that preserves the order of the unobserved resistance ranks without imposing a specific functional form. The estimation proceeds by ordering the observations such that the estimated propensity scores are non-decreasing, $\hat{P}_{(1)} \leq \hat{P}_{(2)} \leq \cdots \leq \hat{P}_{(n_g)}$, where n_g is the number of individuals in group g. The isotonic solver then finds the best-fitting monotone sequence for the corresponding group-specific scores $\hat{P}_i^g$, ensuring that $\hat{u}^g(\cdot)$ satisfies the structural requirement of a quantile-to-quantile map.

The tilt weights, $\omega_g(u)$, are defined as the derivative of the warp function and represent the density ratio between the group-specific and pooled resistance distributions:

$$\widehat{\omega}_g(u) = \frac{\partial \widehat{u}^g(u)}{\partial u}$$

Because the isotonic estimator is piecewise constant, we recover the tilt weights by smoothing the estimated warp function, $\widehat{u}^g(u)$, and taking the derivative of the resulting smooth curve. Specifically, we apply a kernel-weighted local quadratic estimator[20] to the isotonic predictions and define the estimated tilt weights as the implied first derivative. This provides a smooth estimate of how each genetic group is distributed along the latent resistance scale relative to the pooled population. The estimated tilt weights $\widehat{\omega}_g(u)$, multiplied by the sample group shares $\widehat{\pi}_g$, are then normalised so that they sum to one across groups at each value of u, ensuring internal consistency of the mixture identities.

Inference

To account for the fact that the propensity scores, warp functions, tilt weights, and MTEs are all estimated objects, inference is based on a bootstrap procedure in which the full estimation routine is repeated in each replication. In every bootstrap draw, we re-estimate the pooled and group-specific propensity scores, the warp functions, the tilt weights, and the group-specific MTEs. We use 1,000 individual-level bootstrap replications and report original-sample estimates with pointwise 95% normal confidence

[20] We use a Gaussian kernel with bandwidth 0.10.

intervals, calculated as the original-sample estimate plus or minus 1.96 times the bootstrap standard deviation.

3 Empirical Application

3.1 The UK university expansion

Our empirical application exploits the post-war expansion of higher education in the United Kingdom.[21] Although often associated with the Robbins Report of 1963, the expansion was already underway by the late 1950s and early 1960s (Willetts, 2014). Higher education nevertheless remained small and selective: in 1962–63, Great Britain had approximately 216,000 full-time higher-education students, of whom 118,000 attended universities (Robbins, 1963). The Robbins Report endorsed further expansion around the principle that higher education should be available to everyone with the requisite ability and attainment who wished to pursue it.

Expansion proceeded through the creation of new universities, growth at existing institutions, and the development of the non-university sector. The location of new institutions reflected existing provision and practical constraints: the University Grants Committee favoured smaller cities without universities where suitable sites and accommodation were available. The Robbins Report subsequently recommended that further expansion give greater weight to large population centres (Robbins, 1963, pp. 162–163). From the mid-1960s, a binary system distinguished universities from polytechnics, with differences in funding, governance, subjects and admissions (Pratt, 1997). Between 1966 and 1992, university enrolment more than doubled, from approximately 152,000 to 376,000, while enrolment in polytechnics and other colleges increased from approximately 150,000 to 660,000 (Ichino, Rustichini, and Zanella, 2026). The binary divide ended in 1992, when legislative reforms allowed polytechnics to acquire university status (Pratt, 1997).

3.2 Data Sources

3.2.1 UK Biobank

We use individual-level data from the UK Biobank (UKB), a large-scale population-based prospective study (Sudlow et al., 2015). More than 500,000 individuals aged 40 to 69 at recruitment were assessed between 2006 and 2010. Study participants completed a self-administered touch-screen questionnaire, an in-person computer-assisted interview, physical and functional measurements, and provided blood, urine, and saliva samples. The UKB contains information on educational attainment, demographics, and socioeconomic outcomes, although it includes more limited information on individuals' socioeconomic background. However, it provides geographic coordinates for both place of birth and current residence at interview, at a grid resolution of one kilometre, which can be linked to detailed measures of local

[21] Appendix C provides a detailed account of the expansion and its institutional background.

socioeconomic context. In addition, all participants were genotyped, making the UKB one of the largest publicly available genetic datasets.

Our outcome of interest is labour market earnings. Since the UKB does not provide direct information on wages, we proxy earnings by imputing log hourly wages from detailed occupational data, following the approach in Kweon et al. (2020). During the interview, participants answered a series of questions about their job, or their most recent job if recently retired, which allows occupations to be coded into more than 400 Standard Occupation Categories (SOC). We use these occupational data to predict log hourly wages from a parametric model estimated using external wage data from the UK Labour Force Survey (LFS) and the Annual Survey of Hours and Earnings (ASHE). Previous work shows that this occupation-based measure provides a credible proxy for earnings in the UKB (Kweon et al., 2020). We provide further details on the imputation procedure and its reliability in Appendix B.1.

To address potential selection arising from voluntary participation in the UK Biobank, we use inverse probability weights based on UK Census microdata, following van Alten et al. (2024). These weights are designed to make the UKB sample more representative of the underlying population. According to van Alten et al. (2024), the procedure attenuates the associated bias by almost 90%.

3.2.2 Historical District-Level Data

We complement the individual-level UKB data with historical district-level measures from the *Labour Gazette*, the 1951 Census, and the Great Britain Historical Database (Southall and Dorling, 2021). These data characterise the socioeconomic environment of the district of birth before the university expansion and allow us to condition on local characteristics that may be related both to individuals' educational outcomes and to the geography of university openings. In particular, we focus on measures of local labour-market conditions, educational and socioeconomic composition, urban status, population size and density, and demographic growth.

To construct these controls, we combine several sources of district-level historical data. We digitise unemployment rates from the *Labour Gazette*, which provide a measure of local labour-market conditions in the district of birth before or around the period of university expansion. We use the 1951 Census to measure the district's baseline educational, socioeconomic, and urban structure. Educational composition is measured by the share of adult males who remained in full-time education until at least age 15, the statutory school-leaving age at the time.[22] Urban structure is captured by the district's urban/rural classification. Socioeconomic composition is measured using the Registrar General's occupational social-class scheme reported in the 1951 Census. We define high-SES composition as the share of individuals in Social Classes I and II, corresponding to professional, managerial, and related occupations, and low-SES composition as the share in Social Classes IV and V, corresponding to partly skilled and unskilled occupations. Social Class III, which contains the bulk of skilled manual and non-manual

[22]The 1951 Census reports comparable education tabulations for adult males across districts, while corresponding female education tabulations are only available for towns above 50,000 inhabitants. We therefore use male educational composition as a consistent district-level proxy for the share of adults remaining in full-time education until at least age 15, incorporating female measures where comparable data are available.

workers, is left as the omitted middle category. Finally, we use population measures from the Great Britain Historical Database (GBHD) to capture differences in population size, population density, and demographic growth across birth districts.

3.2.3 Higher Education Institutions Data

To measure the changes in geographic access generated by the post-war expansion described in Section 3.1, we construct a historical dataset on higher education institutions (HEI) in the United Kingdom. The dataset is based on the official registry of degree-awarding institutions published by the UK government,[23] which we digitise and enrich with supplementary historical and geographic information. For each institution, we geocode the relevant university location and record key characteristics including institution type, year of establishment, date of university-status recognition, and closure date where applicable. We provide a detailed description of the dataset and its construction in Appendix B.2.

For the purposes of this paper, we treat each geocoded campus providing university-level degree access as a university location. A location enters the university network either when a genuinely new university opens or when an existing institution at that location acquires university status. For simplicity, we refer to both events as university openings. [24] Figure 4 illustrates the resulting spatial and temporal expansion in England and Wales. The panel on the left shows that university openings were geographically widespread but staggered across counties, while the panel on the right shows that the number of active university locations increased from around 85 in 1945 to just over 150 by 1985, with particularly rapid growth during the 1960s. This combination of variation across places and over time allows us to construct cohort-specific changes in university access from individuals' birth locations, which form the basis of the instrument described below. In the robustness analysis, we also distinguish genuinely new university openings from other entries into the university network.

3.3 Measuring Genetic Predisposition to Educational Attainment

To capture genetic predisposition to educational attainment, we use the Educational Attainment Polygenic Index (EA PGI) developed by Lee et al. (2018). Following a now standard approach in the sociogenomic literature (e.g., Barcellos, Carvalho, and Turley, 2018; Barth, Papageorge, and Thom, 2020; Becker et al., 2021; Biroli et al., 2026), we summarise genetic predisposition using a polygenic index rather than individual genetic variants. This reflects the fact that educational attainment, like most outcomes studied in the social sciences, is highly polygenic. That is, there is no single "gene for" education, but rather a very large number of genetic variants, each contributing only a small amount to differences across individuals (Abdellaoui et al., 2023).

A Polygenic Index (PGI)[25] aggregates these small contributions into a single summary measure of

[23] *Source:* GOV.UK guidance on recognised degree-awarding bodies. See Appendix B.2 for further details.

[24] Appendix Figure A.1 provides a more detailed breakdown of university entries by source and entry period, and shows how these different forms of expansion contributed to the growth of the active university network over time.

[25] Also often referred to as a polygenic score (PGS).

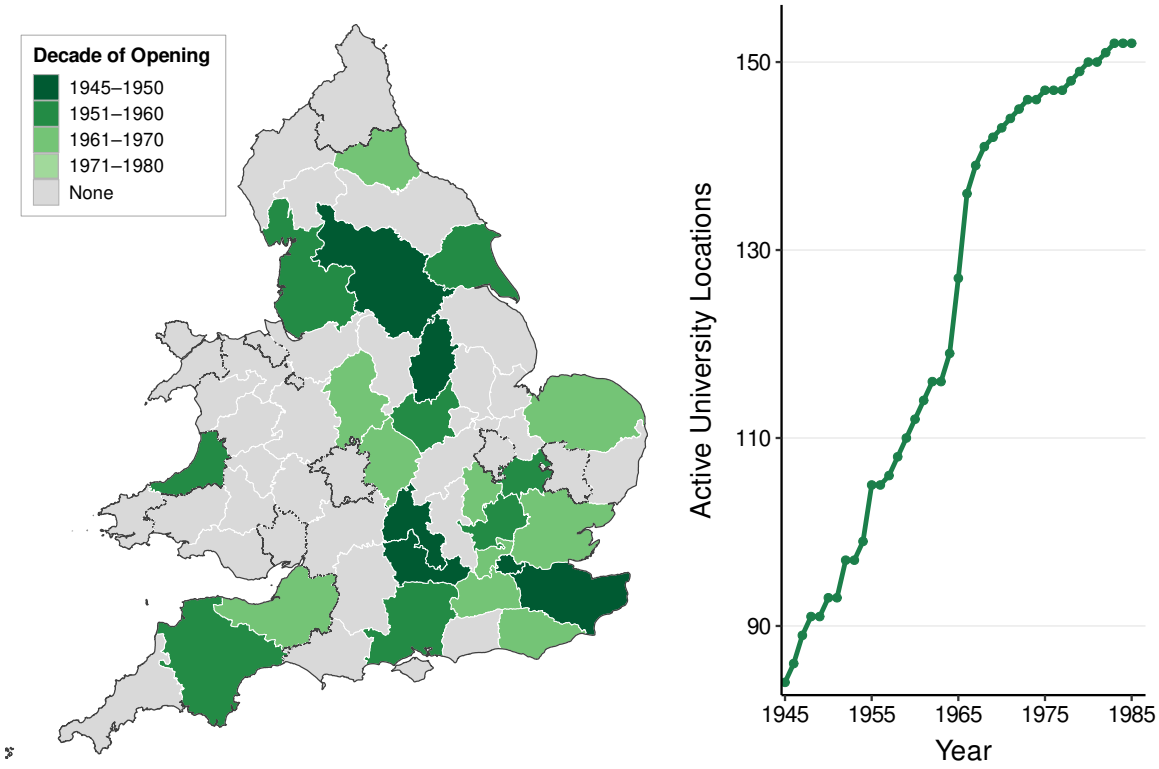


Figure 4: Geographic and Temporal Expansion of University Provision in England and Wales. The figure on the left maps England and Wales at the 1951 county level and shades each county according to the decade in which its first university opening between 1945 and 1980 occurred. Counties in grey experienced no university opening in this period. The figure on the right plots the number of active university locations in the HEI network over time. Source: own calculations from the Higher Education Institutions (HEI) dataset.

an individual's genetic predisposition to a given trait. In the case of educational attainment, the EA PGI combines estimated effects from large-scale genome-wide association studies (GWAS) across a very large number of single nucleotide polymorphisms (SNPs). SNPs are locations in the genome where individuals differ by a single DNA base pair. While humans share the vast majority of their genetic sequence, variation at these sites contributes to systematic differences across individuals. In samples of European ancestry, the EA PGI explains approximately 12–16% of the variation in years of schooling (Okbay et al., 2022).

PGIs are ancestry-specific and are not generally portable across populations, because the predictive relationship between genetic variants and outcomes depends on the demographic and environmental context of the discovery sample, as well as on differences in linkage disequilibrium patterns and allele frequencies across ancestries (Benjamin et al., 2026). Therefore, following standard practice, we restrict the sample to unrelated individuals of European ancestry. Moreover, to adjust for population stratification (Price et al., 2006), we include the first 20 genetic principal components. All genetic data are drawn from the imputed UK Biobank genotype release, and quality control follows Barcellos, Carvalho, and Turley (2021) and Biroli et al. (2026). We construct and standardise the EA PGI to have mean zero and unit

variance in our analytic sample.[26]

Therefore, the EA PGI should be read as a summary measure of an individual's relative genetic predisposition to educational attainment within the relevant ancestry-specific population. It is informative about predisposition to schooling, but it is not a deterministic measure of ability, nor does it identify a single biological mechanism. Rather, it likely captures a broad bundle of small genetic associations related to cognition, non-cognitive traits, preferences, and other characteristics correlated with schooling. In this sense, it is best viewed as a noisy but useful proxy for predisposition to education, rather than as "genes for education" in any literal sense. More generally, the association between a PGI and an outcome should not be interpreted as biologically fixed, since it may vary with the environment (Kweon et al., 2020; Mostafavi et al., 2020).

For the main analysis, consistent with our MTE framework in which genetic predisposition enters as a group-level moderator, we discretise the standardised EA PGI[27] into tertiles (low, medium, and high genetic predisposition). This specification allows us to estimate group-specific marginal treatment effects and facilitates comparisons across levels of genetic predisposition. At the same time, the tertile classification should be interpreted as an econometric grouping device rather than as evidence of three latent genetic types or sharp biological thresholds. As a robustness check, we also estimate specifications that use the continuous PGI directly.

3.4 Reduction in Distance to University

To address the endogeneity of schooling decisions, we compute the reduction in distance to the nearest university during childhood as an instrument for obtaining a university degree. We thus exploit the staggered roll-out of universities to capture plausibly exogenous variation in the cost of accessing higher education. Our approach builds on proximity-based instruments used in the returns-to-schooling literature (e.g. Card, 1995, 1999, 2001; Kling, 2001; Currie and Moretti, 2003; Carneiro, Heckman, and Vytlacil, 2011) and adapts them to the historical expansion of UK higher education institutions in the post-war period. Our strategy also relates to a growing literature exploiting university expansions (e.g. Westphal, Kamhöfer, and Schmitz, 2022; Bratti, Cottini, and Ghinetti, 2022; Berlingieri, Gathmann, and Quinckhardt, 2022; Braccioli et al., 2024; Ichino, Rustichini, and Zanella, 2026).

Using our newly assembled HEI dataset in combination with geocoded birthplace coordinates from the UK Biobank, we compute cohort-specific changes in distance to the nearest university between birth and age 15[28]. Specifically, we define the reduction in distance to the nearest university, $\Delta Dist$, as:

[26] The index is constructed using LDpred (Vilhjálmsson et al., 2015), a Bayesian shrinkage method that accounts for the correlation structure (linkage disequilibrium) across SNPs.

[27] Discretisation of PGIs is common in the literature (e.g., Barcellos, Carvalho, and Turley, 2018; Hollenbach, Schmitz, and Westphal, 2026).

[28] We use age 15 as it is the nearest age to the university enrolment decision that avoids overlap with the Raising of the School Leaving Age (ROSLA). Table A.9 in the appendix shows that the results are robust to alternative exposure windows.

$$\Delta Dist = \delta_{\text{pre}} - \delta_{\text{post}}$$

where δ_{pre} is the distance to the nearest university active in the year of birth, and δ_{post} is the distance to the nearest university active by age 15. Positive values of $\Delta Dist$ indicate that a university opened between birth and age 15 at a distance shorter than the previously nearest institution. A value of zero indicates that changes in the university network between birth and age 15 did not alter the distance to the nearest university.[29] Individuals who experienced a decrease in distance during this period are assumed to have faced a lower cost of enrolment due to the policy-driven expansion of access.

Hence, individuals who experienced a reduction in distance during this period are interpreted as facing a weakly lower cost of accessing university than they would have faced in the absence of the expansion-induced improvement in proximity. This implies that improvements in access should not make university completion less likely. Consistent with this directional response, Appendix Figure A.2 shows that degree attainment rises across bins of the reduction in distance, while Appendix Table A.5 shows a similar pattern after accounting for the baseline controls and fixed effects. Although individual-level monotonicity cannot be tested directly, these patterns are consistent with the monotonicity condition underlying the design.

Since the UK Biobank records birthplace rather than adolescent residence, we use birth location coordinates as a proxy for location in early adolescence. Although this may introduce measurement error for individuals who moved during childhood, available evidence suggests that internal migration in the period of study was arguably low.[30]

3.5 Strategy

As discussed in Section 3.1, the location of new universities was not random. Historical evidence indicates that the expansion partly sought to extend university provision to places with limited existing access, while location decisions also reflected population size, the existing higher-education network, and other local conditions. As a result, exposure to a reduction in distance may be correlated with pre-existing characteristics of an individual's place of birth that also predict degree attainment and later earnings.

To address this concern, we condition on a rich set of predetermined individual and local characteristics. At the individual level, we control for sex and the number of siblings, the latter serving as a proxy for early-life family circumstances,[31] together with 20 genetic principal components to account for ancestry-related population structure (Price et al., 2006). At the local level, we control for baseline university provision, measured by the number of universities active within 10 km of the birthplace at the time of birth, and for a detailed set of characteristics of the district of birth: urban status, educational

[29] Negative values of $\Delta Dist$ arise when the nearest active university at age 15 is farther away than the nearest university active at birth, for example because an institution closed or relocated.

[30] As a benchmark, using the 1961 Census, we approximate that only around 4% of children under 15 relocated across districts. Consistent with this, von Hinke and Vitt (2024) estimate an average annual mobility in the UKB sample of 1.2%.

[31] This follows the approach by Barcellos, Carvalho, and Turley (2021).

composition, log population, log population density, population growth between 1946 and 1951, the shares of high- and low-SES residents, and the historical unemployment rate.[32] Finally, our preferred specification includes birth-region-by-cohort fixed effects, which absorb persistent differences across regions, common differences across cohorts, and cohort-specific shocks that vary across regions.

Because our instrument is defined by geographic distance, there is a trade-off in how finely geographic heterogeneity should be absorbed. Our specification combines birth-region-by-cohort fixed effects with the predetermined local characteristics described above rather than using substantially finer county-by-cohort fixed effects. Increasingly granular geographic fixed effects can absorb much of the spatial variation that generates a distance-based instrument, thereby reducing identifying variation and statistical power. Becker, Tommasi, and Zhang (2025) emphasise this trade-off and show that, when economically relevant local confounders are observed, conditioning directly on them can preserve considerably more identifying variation than relying on increasingly fine geographic fixed effects.

Conditional on these predetermined characteristics and fixed effects, we use the remaining variation in $\Delta Dist$ as our source of quasi-experimental variation in access to higher education. Section 4.2 examines the credibility of this identifying variation directly, including balance with predetermined characteristics and the EA PGI, robustness to substantially finer county-by-cohort fixed effects and alternative instrument specifications, and a series of timing-placebo exercises.

3.6 Sample and Descriptive Statistics

Our analysis focuses on UKB participants born between 1945 and 1969 in England and Wales, whose childhood years overlap with the post-war expansion of the university system.[33] We further restrict the sample to individuals in the European-ancestry genetic sample and require non-missing information for the variables entering our baseline specifications. These restrictions yield a final analysis sample of 157,109 individuals.

We report descriptive statistics for the main variables in our analysis in Table 1. Just over one third of individuals in the estimation sample are university graduates, and the mean log hourly wage is 2.654, corresponding to an hourly wage of about £14.21. The university expansion generated substantial variation in access: the distance to the nearest university fell by 3.85 km on average between birth and age 15, and 29.1 percent of individuals experienced a strictly positive reduction in distance. Among those experiencing a positive reduction, the average decrease in distance is approximately 13.28 km. [34]

[32] These characteristics are measured before, or at the beginning of, the post-war university expansion wherever possible. Following a similar logic to Abramitzky, Boustan, and Giuntella (2025), we fix baseline district characteristics before the relevant expansion so as to avoid conditioning on local outcomes that may themselves have responded to university entry. Urban status, educational and socioeconomic composition, population, and population density are measured in 1951, while population growth is computed over 1946–1951. Historical unemployment is constructed from the *Labour Gazette*. See Section 3.2.2 and Appendix Table A.1 for detailed definitions and sources.

[33] Scotland is excluded because comparable pre-expansion district-level covariates are not consistently available for Scottish birth locations. In addition, the Scottish education and university systems differed institutionally from those of England and Wales during the period studied, limiting comparability.

[34] Appendix Table A.3 provides additional descriptive statistics for the instrument. A broader set of summary statistics for the analysis sample and the covariates used in the empirical specifications is reported in Appendix Table A.2, while Appendix

Table 1: Descriptive Statistics by Graduate Status

	(1) Mean	(2) Non-Graduates	(3) University Graduates	(4) Difference (3) − (2)
University degree	0.361			
	(0.480)			
Log hourly wage	2.654	2.559	2.821	0.262
	(0.345)	(0.335)	(0.293)	(0.003)
EA PGI	-0.018	-0.197	0.297	0.494
	(0.986)	(0.950)	(0.969)	(0.009)
$\Delta Dist$	3.848	3.492	4.477	0.985
	(11.316)	(10.605)	(12.447)	(0.177)
$\mathbb{I}\{\Delta Dist > 0\}$	0.291	0.282	0.305	0.023
	(0.454)	(0.450)	(0.461)	(0.008)
N	157,109	94,878	62,231	157,109

Notes: Column (1) reports the weighted pooled mean. Columns (2) and (3) divide individuals according to whether they hold a university degree. Standard deviations are reported in parentheses beneath these means. Column (4) reports the mean among university graduates minus the mean among non-graduates, with district-of-birth-clustered robust standard errors. Means, standard deviations and contrasts use UKB weights. Reduction in distance is measured in kilometres. $\mathbb{I}\{\Delta Dist > 0\}$ is an indicator equal to one for a strictly positive reduction, so its mean is the share of individuals who experienced one. The unit of observation is the individual.

The differences by graduate status are substantial. University graduates earn 0.262 log points more per hour and score almost half a standard deviation higher on the EA PGI than non-graduates. They also experienced somewhat larger improvements in university access: the average reduction in distance is 4.48 km among university graduates, compared with 3.49 km among non-graduates, and they are 2.3 percentage points more likely to experience any reduction at all. The strong association between degree attainment and the EA PGI points to substantial selection into higher education, while the difference in exposure to university openings is consistent with the first-stage relationship that we exploit below.

To explore whether the raw data already show signs of a negative gene–environment interaction, we examine how earnings vary jointly with degree attainment and the EA PGI. Figure 5 plots average log hourly wages across the EA-PGI distribution separately for university graduates and non-graduates. Wages rise with the EA PGI in both groups, while graduates earn substantially more throughout the distribution. Importantly, the unadjusted wage gap between graduates and non-graduates is largest at lower levels of the EA PGI and narrows as the PGI increases. This descriptive pattern is consistent with a negative $G \times E$, whereby university may generate larger earnings gains among individuals with a lower genetic predisposition towards educational attainment, and is therefore suggestive of a compensatory role for higher education.

Table A.1 provides detailed variable definitions.

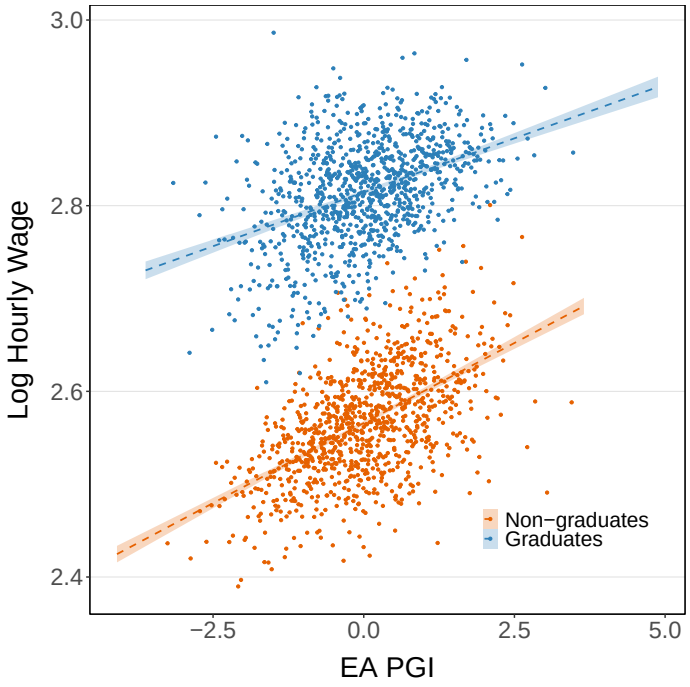


Figure 5: Log Hourly Wages by EA PGI and Graduate Status. Points show averages of imputed log hourly wages (Kweon et al., 2020) within 1,000 EA PGI bins, separately for university graduates and non-graduates. Lines show linear fits of log hourly wages on EA PGI, estimated separately by graduate status on the full sample, using UKB weights (van Alten et al., 2024). Shaded bands indicate 95% confidence intervals.

4 Results

4.1 First Stage and Selection into University

We begin by examining the first stage and the patterns of selection into university that underlie the MTE analysis. As described in Section 2.6, we estimate both pooled, $P(Z)$, and EA PGI-group-conditional propensity scores, $P^g(Z)$, using Probit models. We report the estimated average marginal effects of the reduction in distance to the nearest university in Table 2.

Improved university access predicts degree attainment both in the pooled sample and within each EA PGI tertile. A 10 km reduction in distance increases the probability of degree attainment by 1.4 percentage points overall. The corresponding marginal effects are 1.2, 1.5, and 1.2 percentage points in the low, middle, and high EA PGI tertiles, respectively. Moreover, a Wald F-test does not reject equality of the marginal effects across the three groups ($p = 0.374$), suggesting that the first-stage response to improved university access is broadly similar across the EA PGI distribution.

Despite this similarity in responsiveness to university access, selection into university differs sharply across the EA PGI distribution. The proportion of individuals holding a university degree rises from 24.5 percent in the lowest EA PGI tertile to 35.0 percent in the middle tertile and 50.1 percent in the highest.

Table 2: First-Stage Effects of University Access on University Attainment by EA PGI Tertile

	(1) Overall	(2) Low EA PGI	(3) Medium EA PGI	(4) High EA PGI
$\Delta Dist$	0.014^{***}	0.012^{***}	0.015^{***}	0.012^{***}
	(0.003)	(0.002)	(0.003)	(0.003)
Mean degree attainment	0.361	0.245	0.350	0.501
Observations	157,109	52,370	52,369	52,370
Controls	Yes	Yes	Yes	Yes
Region × cohort FE	Yes	Yes	Yes	Yes
Joint test: equal first-stage AMEs across PGI tertiles	—		$p = 0.374$	

Notes: Entries report average marginal effects (AMEs) from Probit models estimated with UK Biobank inverse-probability weights; AMEs and outcome means are also averaged using those weights. The Overall column comes from the baseline pooled model, which does not control for EA PGI. The tertile columns come from a pooled model containing EA PGI-tertile indicators and interactions between the tertiles and the reduction in distance. Reduction in distance, $\Delta Dist$, is measured per 10 km. All specifications include the baseline controls and birth-region-by-cohort fixed effects. Delta-method standard errors, clustered by numerical birth-district ID, are reported in parentheses. Inference uses birth-district cluster degrees of freedom. The joint row reports the two-restriction Wald F-test that the three first-stage AMEs are equal. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hence, the main difference across genetic groups is not in how strongly degree attainment responds to the expansion, but in how the groups are distributed across the propensity to hold a university degree.

Figure 6 makes these differences visible in the estimated propensity-score distributions. The left panel plots the pooled propensity score, $\widehat{P}(Z)$, while the right panel shows the corresponding group-specific propensity scores, $\widehat{P}^{g}(Z)$, obtained by estimating the same model separately within EA PGI tertiles. The pooled propensity score spans a broad range, indicating substantial variation in the probability of degree attainment generated by the first stage. There is also considerable overlap across EA PGI groups, although their distributions are clearly shifted and span somewhat different ranges of support. Individuals in the high EA PGI tertile are concentrated at higher propensities to hold a university degree, while those in the low tertile are concentrated at lower propensities. The separation across groups therefore highlights the strong relationship between EA PGI and selection into university.

4.2 Validity of the Instrument

We next assess the validity of our identifying assumptions and the robustness of the first stage. Specifically, we examine threats arising from the non-random placement of universities, unobserved local trajectories and direct local effects of university openings, imbalance in predetermined characteristics, and sensitivity to alternative instrument constructions.

We begin by asking whether university openings were systematically related to pre-existing local conditions. Appendix Table A.4 relates university entry between 1945 and 1980 to the district characteristics included in our baseline control set. Population size and existing university provision are the

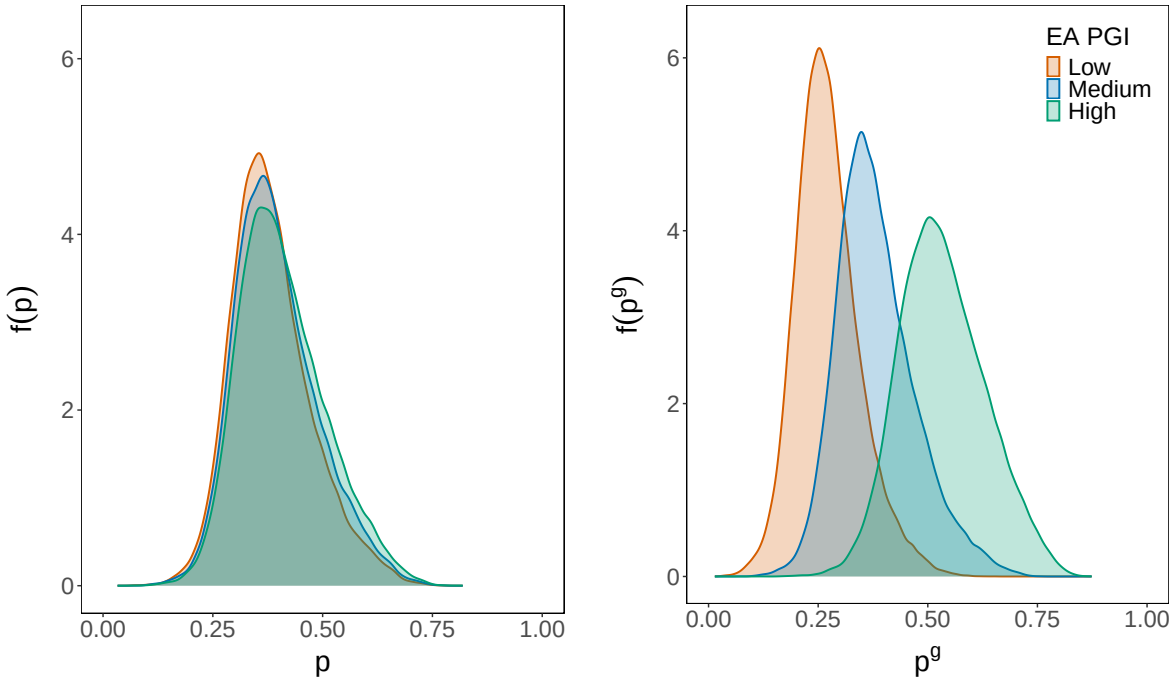


Figure 6: Propensity Scores of Obtaining a University Degree by EA PGI Tertile. The left panel shows the distribution, by EA PGI tertile, of the estimated unconditional propensity score, $P(Z)$, from a Probit model of university degree estimated on the whole sample. The right panel shows the estimated conditional (group-specific) propensity scores, $P^g(Z)$, from separate Probit models estimated within each EA PGI tertile. All models include the baseline controls.

most consistent predictors of subsequent openings, with some evidence of relationships with historical unemployment and educational composition, while the remaining characteristics do not systematically predict entry. These results confirm that university placement was not random, but suggest that observable targeting was concentrated along a limited set of dimensions, rather than spanning across a wide range of pre-expansion socioeconomic characteristics.

These placement regressions, however, cannot rule out unobserved local conditions or cohort-specific trajectories within the broad regions used in our baseline specification. We therefore examine in Appendix Table A.6 whether residual local confounding matters by replacing birth-region-by-cohort fixed effects with substantially finer birth-county-by-cohort fixed effects.[35] This more demanding specification absorbs all between-county variation in university access within a birth cohort and, as expected for a distance-based instrument, reduces precision.[36] Nevertheless, the point estimate remains statistically indistinguishable to the baseline estimate, providing reassurance that unobserved geographic confounding within broad regions is unlikely to explain the first-stage relationship.

Additionally, our identifying assumptions require that the instrument should be unrelated to predeter-

[35] Our harmonised geography comprises 11 broad regions and 62 counties in England and Wales. The first stage is also robust to alternative region-level fixed-effect specifications.

[36] The first-stage F-statistic falls from 35.47 to 10.78.

mined individual characteristics conditional on the baseline specification. This is particularly important for the EA PGI, since Assumption (A3) requires the instrument to be independent of the group-defining variable G.[37] Appendix Table A.8 examines this directly. We find no significant relationship between the reduction in distance and the EA PGI, with an estimated coefficient of less than 0.01 standard deviations per 10 km reduction in distance. Moreover, we find similar balance across the other predetermined outcomes.

A further concern is that the first-stage relationship could reflect persistent differences across places and cohorts rather than changes in university access during childhood, while university openings could themselves affect local economic conditions and hence later earnings. To assess these possibilities, Appendix Table A.7 reconstructs the reduction in distance using openings occurring only after age 25, when the focal individual's decision about whether to attend university had already been made. Reassuringly, these placebo measures predict neither university completion nor log hourly wages, suggesting that neither persistent place–cohort differences nor direct local effects of university expansion are likely to explain the main relationships.[38]

Finally, we examine whether the results depend on how our instrument is constructed. Appendix Table A.9 shows that the first-stage and reduced-form relationships remain similar across alternative childhood exposure windows and when we distinguish genuinely new university openings from other entries into the university network. Appendix Table A.10 further shows that the results are robust to alternative functional forms of geographic access. Together, these exercises indicate that the empirical relationships are not driven by a particular choice of exposure window, definition of university entry, or functional form of the distance measure.

4.3 Rank Comparability

Section 4.1 showed that selection into university differs sharply across EA PGI tertiles. This raises the question of whether the pooled resistance rank U and the group-specific ranks U^g are directly comparable across groups. Under distributional invariance of unobserved resistance, the mapping between them is the identity, $u^g(u) = u$ for all $u \in [0, 1]$. To assess whether this condition holds, we estimate the warp functions $u^g(\cdot)$ using the calibration identity described in Section 2.6.

The left panel of Figure 7 reports the estimated warp functions $\hat{u}^g(u)$ for the three EA PGI tertiles plotted against the pooled resistance rank u. The estimates display a clear and monotonic ordering that mirrors the selection patterns observed in Figure 6. In particular, the warp for the High EA PGI group lies consistently above the 45-degree line ($u^H(u) > u$). This reflects the fact that High EA PGI individuals are disproportionately concentrated at low resistance (high propensity) levels in the pooled distribution. As a consequence, a given pooled resistance rank corresponds to a higher percentile of resistance within the High EA PGI group. For example, an individual at the median of the pooled

[37] In the theoretical exposition, we suppress the conditioning variables X for notational simplicity. In the empirical application, the corresponding requirement is independent conditional on X.

[38] The placebo exercise considers several overlapping post-entry exposure windows. Joint tests of the corresponding coefficients yield $p = 0.843$ for degree attainment and $p = 0.831$ for wages.

population ($u = 0.5$) maps to approximately the 62nd percentile of the resistance distribution within the High EA PGI group ($u^H \approx 0.62$). This means that because High EA PGI individuals are, on average, less resistant to university, a person with median pooled resistance is relatively resistant compared to most High EA PGI peers. Conversely, the warp for the Low EA PGI group lies below the 45-degree line ($u^L(u) < u$), indicating that individuals in this group are more concentrated at higher resistance levels. Thus, a pooled-median individual corresponds to a relatively low percentile of resistance within the Low EA PGI distribution. The Medium EA PGI group tracks the 45-degree line more closely, consistent with its central position in the pooled population.

These patterns imply that the distribution of unobserved resistance is not identical between EA PGI groups ($u^g(u) \neq u$). The systematic deviation of the estimated warp functions from the 45-degree line shows that the same pooled rank corresponds to different levels of latent resistance between groups. In other words, individuals located at the same percentile of their group-specific distribution do not face the same underlying cost of attending university when conditioning on EA PGI. This has direct implications for the interpretation of group-specific MTE curves. If we were to compare MTEs at the same native rank within each group, we would be comparing individuals facing different latent resistance levels rather than individuals at the same margin of indifference. For instance, a pooled-median individual ($u = 0.5$) maps to, approximately, $u^H(0.5) \approx 0.62$ in the High EA PGI group and to $u^L(0.5) \approx 0.36$ in the Low EA PGI group, indicating that the same latent resistance corresponds to different percentile positions across groups. To ensure comparability, we therefore use the estimated warp functions to re-express group-specific MTEs on the common pooled resistance scale u, so that effects are evaluated at the same underlying level of resistance to university.

We then display in the right panel of Figure 7 the estimated tilt weights, $\omega_g(u) = \hat{u}^{g\prime}(u)$, obtained by smoothing the derivative of the estimated warp (see Section 2.6). The plotted mixture weights include the group shares and are normalised to sum to one across groups. They describe how densely individuals from each EA PGI group populate different parts of the pooled resistance distribution. As expected, the High EA PGI tilt weight declines with u, indicating over-representation at low resistance levels and under-representation at high resistance levels. Conversely, the Low EA PGI tilt weight increases with u, reflecting concentration among high-resistance individuals. The Medium EA PGI group exhibits comparatively stable weights across the distribution, consistent with its central position in the pooled sample.

Reassuringly, the patterns documented above are not particularly sensitive to how the warp and tilt functions are estimated. Appendix Table A.14 shows that replacing the baseline probit with a logit or linear probability model leaves the estimated warp functions virtually unchanged, with a maximum absolute deviation of just 0.011. The tilt weights display similarly stable profiles across alternative smoothing bandwidths within the common support, with noticeable departures only for the smallest bandwidths (Appendix Figure A.3).

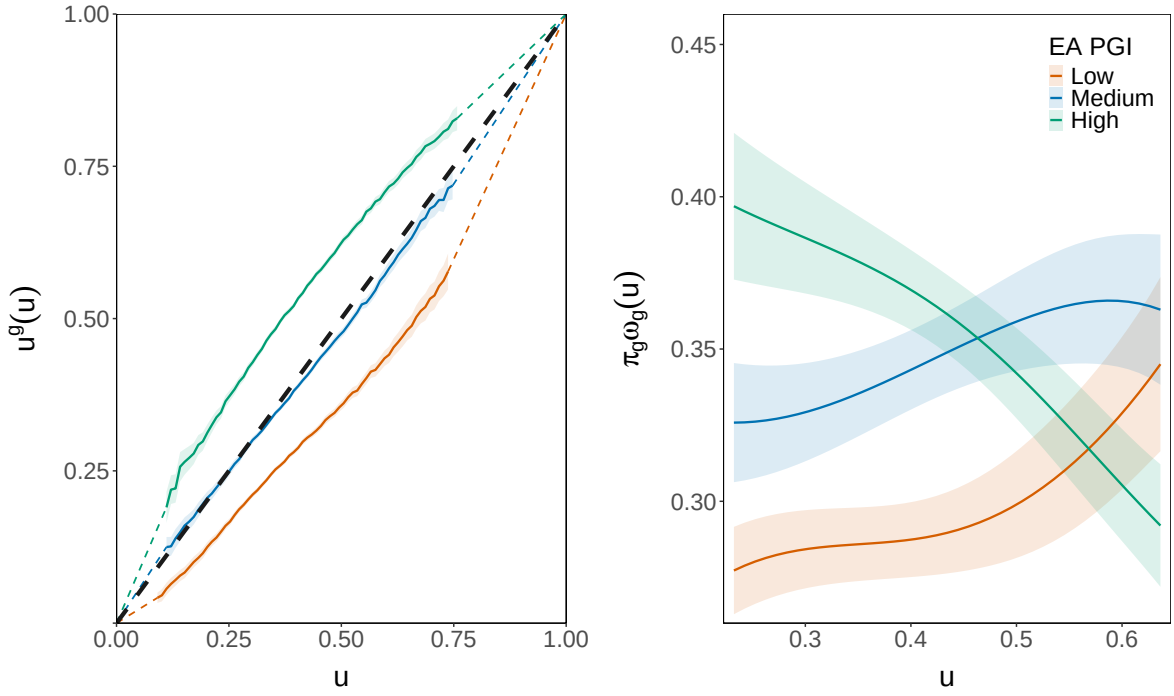


Figure 7: Estimated Warp Functions and Mixture Weights. The left panel shows the estimated warp functions, $\hat{u}^g(u)$, which map the pooled resistance rank u into each group's resistance rank, with the identity line for reference. The right panel shows the estimated mixture weights, $\hat{\pi}_g\,\hat{\omega}_g(u)$: the group share times the tilt weight $\omega_g(u) = (u^g)'(u)$, the derivative of the warp (Proposition 1), that is, the probability of belonging to group g at pooled rank u, which sums to one across groups. Curves are original-sample estimates, and shaded areas are pointwise 95% normal intervals from 1,000 bootstrap replications. The mixture-weight panel is restricted to the interval estimated in all bootstrap draws. Dashed coloured warp segments connect the supported estimates to the theoretical endpoints and indicate extrapolation; they are not used for the common-support treatment effects. The tilt weights are estimated by smoothing the derivative of the estimated warp with bandwidth 0.10.

4.4 Estimation of Marginal Returns to Education

4.4.1 Marginal Treatment Effects in the Pooled Population

We begin by presenting the estimated marginal treatment effects for the pooled population. MTEs are properly identified only over the region of the propensity score where both treated and untreated observations are present and where the instrument generates sufficient local variation. To ensure reliable local IV estimates, we therefore restrict attention to the common support of the estimated propensity score and trim the lower and upper tails by 0.5%. This procedure yields a common support interval of $u \in [0.211,\ 0.659]$.[39]

Figure 8 plots the estimated pooled MTE curve over the available region with pointwise 95% normal confidence bands computed from the bootstrap[40]. The estimates reveal a clear downward-sloping pat-

[39]Table A.12 in the Appendix presents the estimated common support ranges across different specifications and trimming rules. For plotting, we additionally require finite estimates in all 1,000 bootstrap draws, leaving approximately $u \in [0.220, 0.645]$.

[40]We use 1000 bootstrap replications since they provide a better approximation to standard errors than asymptotic standard

tern: individuals with low resistance to completing university exhibit substantially higher returns, while marginal returns decline as resistance increases. The downward-sloping pattern of marginal returns is also robust to alternative smoothing bandwidths. Appendix Figure A.4 compares the pooled MTE estimated with bandwidths from 0.05 to 0.20, holding all other choices fixed. The alternative estimates closely track the baseline, preserving the same overall pattern of declining returns across the identified support. In our chosen specification, estimated returns decline from approximately 0.90 log points at the lower displayed boundary to 0.11 log points at the upper boundary. This pattern is consistent with the widely documented evidence of positive selection on gains (e.g. Carneiro, Heckman, and Vytlacil, 2011; Kamhöfer, Schmitz, and Westphal, 2019), whereby individuals who are less resistant to completing university also tend to experience larger gains.

Panel A of Table 3 reports the pooled treatment effects implied by the MTE. Averaging over the trimmed common support, we obtain an average estimated return to a university degree of approximately 0.491 log points. This magnitude is close to the benchmark 2SLS estimate reported in Appendix Table A.13 and is broadly comparable in magnitude to the sizeable graduate wage differentials documented in the United Kingdom.[41] The ATT is 0.718 and the ATU is 0.332. The positive ATU suggests that non-attendance need not imply an absence of wage gains from university, leaving scope for policies that reduce barriers to completion to benefit some non-graduates. However, the wide confidence interval and the exclusion of the highest-resistance individuals from the estimated support limit what we can conclude about the gains from extending access to those individuals.

Because the baseline support is limited, its reported ATE averages returns over the identified region and only approximates the full-population ATE. To expand the range over which we can estimate returns, we include continuous EA PGI as an additional covariate in both the selection and outcome equations. As the preceding subsections show, genetic predisposition is an important predictor of university completion. Including it in X broadens the propensity-score distribution. We evaluate the resulting pooled MTE over the wider interval $u \in [0.06, 0.90]$, with the corresponding curve reported in Appendix G.2. Under the maintained model assumptions, both specifications would identify the same population ATE if full support were available. Here, however, they average returns over different parts of the resistance distribution, so we would not expect the estimates to coincide exactly. Reassuringly, the wider-support estimate of 0.437 log points is close to the baseline estimate of 0.491. This similarity is consistent with recovering comparable average returns whether genetic predisposition enters through the unobserved determinants of selection and outcomes or is explicitly included among the observed covariates.[42]

Overall, the pooled MTE reveals economically meaningful heterogeneity in the returns to a university degree along the resistance margin and a large wage premium for university graduates.

errors (Heckman, Ichimura, and Todd, 1997). Throughout, we re-estimate objects $P(Z), P^g(Z), u^g(u), w^g(u)$ in each iteration of the bootstrap so that all standard errors account for the fact that these objects are estimated objects themselves.

[41] Blundell, Green, and Jin (2016), using the UK Labour Force Survey, document the age profile of the median hourly-wage gap between university graduates and school-leavers. For cohorts overlapping closely with our sample, graduate wages remain substantially higher through the 40s and 50s. These estimates are descriptive wage differentials rather than causal returns to a university degree and are therefore not directly comparable to our MTE estimates.

[42] We leave G outside X in the baseline to retain a pooled resistance rank whose distribution can differ across genetic groups. Conditioning on G instead defines resistance ranks conditional on genetic predisposition, rather than the common pooled rank used in our group comparisons.

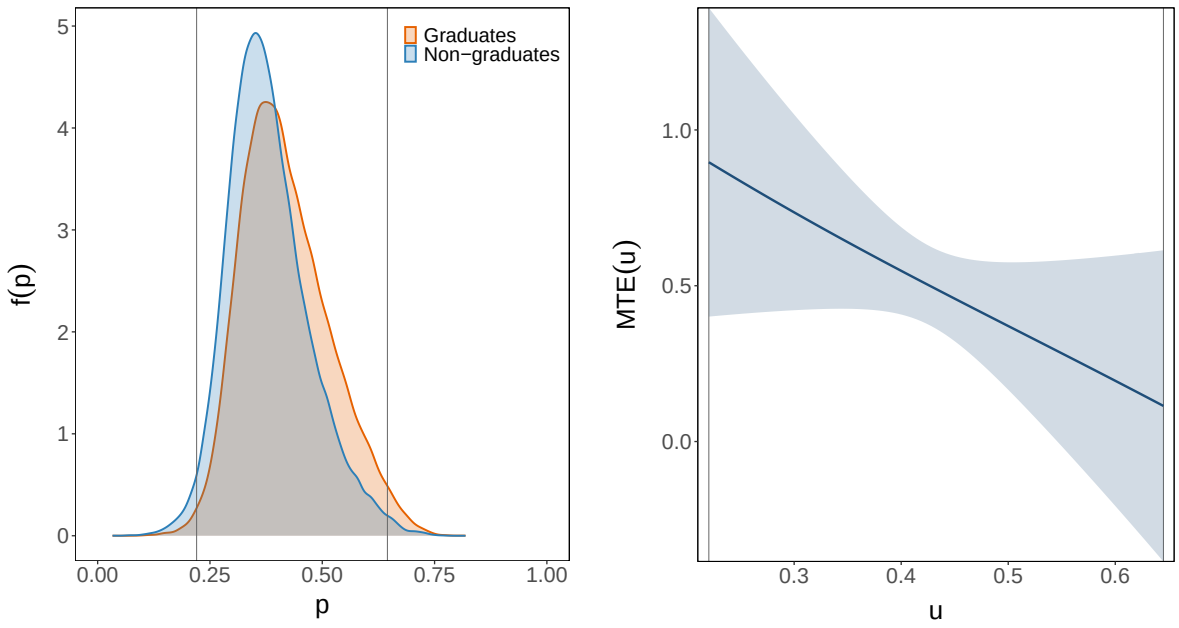


Figure 8: Propensity Score and Pooled MTE. The left panel shows the estimated propensity-score densities for graduates and non-graduates; vertical lines mark the pooled support estimated in all bootstrap draws, approximately $[0.220, 0.645]$. The right panel shows the original-sample pooled MTE with pointwise 95% normal intervals from 1,000 bootstrap replications. Each draw trims 0.5% of each tail and restricts the support to the overlap between treated and untreated individuals. The smoothing parameters are described in Section 2.6.

4.4.2 Group-Specific Marginal Returns to Education

We now examine how marginal returns to a university degree vary across EA PGI groups. Using the trimmed treated/untreated overlap restrictions and the warp mappings, we compare the curves over the common pooled-resistance interval $u \in [0.232, 0.636]$.

The left panel of Figure 9 displays the estimated group-specific MTE curves projected onto the common resistance scale u using the warp transformation. Each curve is evaluated at the group-specific sample mean covariate vector $\bar{X}_g$, so that the plotted objects are the same ones used to reconstruct the pooled MTE in the right panel. Two patterns emerge. First, there are substantial differences in levels across EA PGI groups. Individuals in the high EA PGI group exhibit consistently lower marginal returns throughout the common support, whereas individuals in the low and middle groups experience considerably larger gains. Second, the MTE curves differ not only in levels but also in shape. In particular, the middle EA PGI group displays a steeper decline in returns as resistance increases, while the low and high groups exhibit flatter and quasi parallel profiles. These findings indicate that heterogeneity in returns is not monotonic in genetic predisposition and varies across the distribution of resistance. However, confidence intervals are wide, particularly near the boundaries, and differences between groups are estimated imprecisely.

We then report the corresponding restricted-support treatment effects in Panel B of Table 3. In the original sample, the common pooled interval maps into group-specific intervals $U^L \in [0.140, 0.481]$,

Table 3: Returns to University Education

	Resistance Support				Treatment Effects†		
	$u^g_{\min}$	$u^g_{\max}$	$u_{\min}$	$u_{\max}$	ATE	ATT	ATU
Panel A: Pooled MTE							
Unconditional			0.211	0.659	0.491***	0.718***	0.332**
					(0.068)	(0.159)	(0.132)
Continuous G in X			0.060	0.900	0.437***	0.911***	0.119
					(0.091)	(0.223)	(0.210)
Panel B: Group-Conditional MTE							
Low EA PGI	0.140	0.481	0.222	0.646	0.687***	0.742**	0.646**
					(0.175)	(0.337)	(0.281)
Middle EA PGI	0.223	0.623	0.222	0.646	0.543***	0.796***	0.374*
					(0.112)	(0.239)	(0.193)
High EA PGI	0.338	0.747	0.222	0.646	0.329***	0.430	0.239
					(0.120)	(0.278)	(0.239)
Panel C: Aggregation of Group Effects							
Unconditional (common support)			0.222	0.646	0.492***	0.709***	0.340***
					(0.067)	(0.155)	(0.127)
Aggregated group effects			0.222	0.646	0.519***	0.656***	0.420***
					(0.080)	(0.168)	(0.140)
Direct minus aggregated					-0.027	0.054	-0.080
					(0.036)	(0.070)	(0.067)

Notes: Estimated treatment-effect parameters across MTE specifications. Entries are original-sample estimates in log-wage points, with bootstrap standard errors in parentheses from 1,000 individual-resampled replications per specification. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ from two-sided normal tests using the bootstrap standard deviation. Estimation is unweighted, with additive birth-region and cohort fixed effects. The unconditional and group-specific specifications use a LOESS span of 0.10 in the covariate step and a bandwidth of 0.15 in the derivative step, with 0.5% trimming per tail; the specification with continuous G in X uses 0.04 and 0.10 and the fixed reporting interval $[0.06, 0.90]$. Support columns give the original-sample integration limits. Panel C aggregates the group effects using sample group shares rather than integrating the reconstructed curve; its differences use paired bootstrap draws. † ATE, ATT and ATU denote the average treatment effect, the average effect among graduates and the average effect among non-graduates, respectively. They are evaluated over the observed common support of the resistance distribution and are therefore restricted-support parameters rather than full-population effects over $[0, 1]$.

$U^M \in [0.223, 0.623]$ and $U^H \in [0.338, 0.747]$ for the low, middle and high EA PGI groups. The estimated ATEs are 0.687, 0.543 and 0.329 log points, respectively. These point estimates suggest larger average returns among individuals with lower genetic predisposition to education, but the cross-group contrasts are substantially less precise than the individual group averages. Moreover, the group-specific MTEs display a similar stability across alternative smoothing bandwidths. As shown in Appendix Figure A.5, the broad differences in returns across EA PGI groups are preserved, with the alternative curves remaining close to the baseline estimates over most of the common support. The implied treatment effects are reported in Appendix Table A.16.

Finally, we assess the empirical reconstruction implied by Corollary 1. Specifically, we map each group-specific MTE onto the pooled resistance scale and aggregate the curves using the estimated tilt weights and group shares. The right panel of Figure 9 compares this reconstruction with the directly estimated pooled

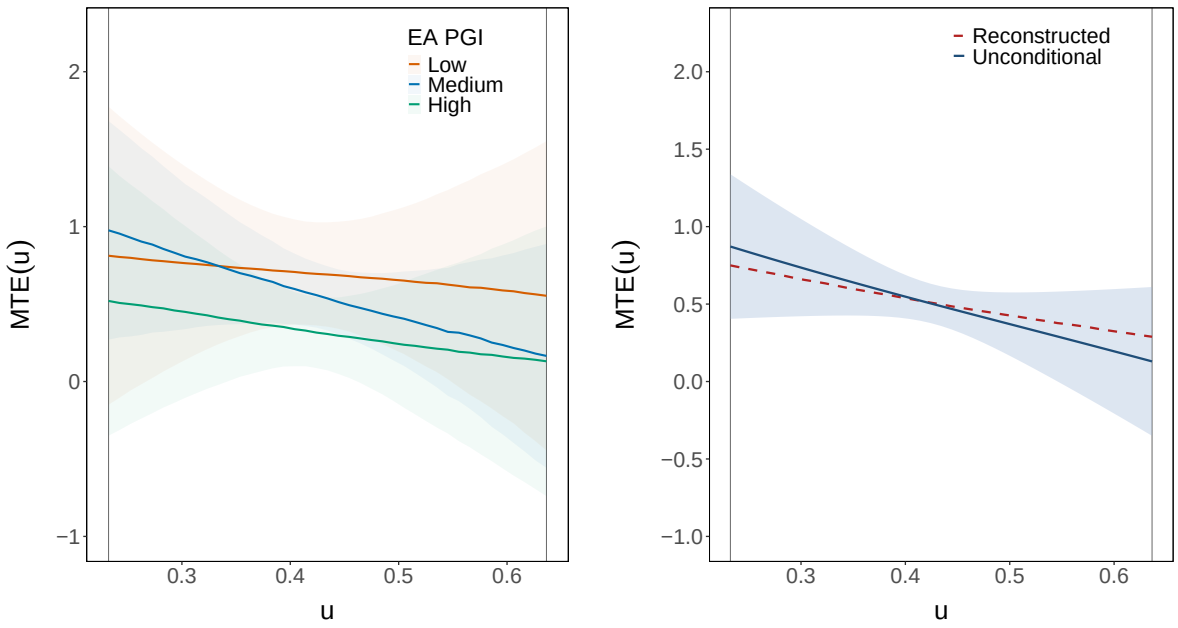


Figure 9: Group-Specific MTEs and Pooled Reconstruction. The left panel shows the group-specific MTEs, evaluated at each group's mean covariates and mapped onto the pooled resistance scale. The right panel compares the unconditional pooled MTE estimated directly (solid) with the pooled MTE reconstructed from the group-specific curves using the derivative-based mixture weights, as in Corollary 1 (dashed). Curves are original-sample estimates, and shaded areas are pointwise 95% normal intervals from 1,000 bootstrap replications; in the right panel, the interval shown is that of the directly estimated unconditional curve. Vertical lines mark the common support estimated in all bootstrap draws, approximately $[0.232, 0.636]$. The smoothing parameters are described in Section 2.6.

MTE over the same common support. The curves lie close together, particularly in the interior. Over the interval estimated in all bootstrap draws ($u \in [0.232, 0.636]$), the reconstructed curve exceeds the direct estimate by 0.015 log points on average (95% normal confidence interval $[-0.044, 0.073]$), indicating no statistically detectable average discrepancy. Panel C of Table 3 likewise reports no statistically significant differences at the 5% level between direct pooled treatment effects and aggregates of the group-specific effects. Together, these comparisons are reassuring about the reconstruction, although agreement is weaker near the less precisely estimated tails.

4.5 Decomposition of Gene-Environment Interplay

We now quantify the extent to which differences in educational returns across EA PGI groups reflect differences in selection into university (rGE) versus differences in returns conditional on resistance ($G \times E$). Following Proposition 3, we decompose cross-group differences in treatment effects into three components. First, the returns (MTE) channel, which captures differences in marginal returns at a common resistance margin and therefore maps naturally into $G \times E$. Second, the target-weight channel, and third, the composition channel. These latter two capture two ways in which the same marginal gains

may be aggregated differently across groups: the former through the estimand-specific weights placed on each resistance margin, and the latter through where each group is located along the pooled resistance scale. For treatment parameters whose weights depend only on propensity-score objects, such as the ATE, ATT, and ATU, we interpret these latter two terms jointly as a selection channel, that is, as active rGE.

Table A.15 and Figure 10 report the decomposition. For the ATE, the target-weight channel is zero by construction, and we report the two-channel decomposition into returns and composition. The high–low ATE gap is −0.355 log points, comprising a returns component of −0.371 and an offsetting composition component of 0.015. The total gap and returns component are significant at the 10% level, but not at 5%; the composition component is imprecise. The high–middle gap is −0.216, with a returns component of −0.241 and composition of 0.025. The middle–low gap is −0.140, comprising −0.143 in returns and 0.003 in composition. These estimates suggest that returns account for most of the estimated gaps, while leaving substantial uncertainty about their magnitude.

For the ATT and ATU we retain all three channels. In the high–low ATT comparison, the returns component is −0.303, the target component −0.088 and the composition component 0.079. For the ATU, the corresponding estimates are −0.413, 0.037 and −0.032. The selection-related components largely offset in these point estimates, but the total differences and returns components have wide confidence intervals. This pattern is robust to the choice of smoothing bandwidth (Appendix Table A.17): across bandwidths, the returns component remains the main source of the cross-group differences, while the selection-related components remain small and largely offsetting.

Overall, the point estimates attribute most of the cross-group differences to returns at comparable resistance margins rather than selection-related weighting. This is suggestive evidence of a compensatory role for university education, not a precise rejection of economically relevant sorting effects.

5 Conclusions

In this paper, we develop a framework for comparing treatment effects across groups when group identity shifts the distribution of latent heterogeneity relevant for both selection and outcomes. Our main methodological point is that, in a Marginal Treatment Effect (MTE) setting, group-specific MTEs are not directly comparable when groups differ in their distributions of latent resistance to treatment. In such cases, standard subgroup comparisons are only quantile-to-quantile comparisons within each group and need not correspond to comparisons at a common underlying margin of resistance. We address this problem by introducing a monotone warp function that maps group-specific resistance ranks onto a common pooled scale. This alignment delivers a common metric for comparing marginal returns across groups and implies that the pooled MTE can be recovered as a weighted combination of the aligned group-specific MTEs.

Building on this alignment, we show that between-group differences in treatment-effect parameters can be decomposed into a returns component and selection-related components. In our application, this distinction maps naturally into gene–environment interaction (G×E), on the one hand, and active

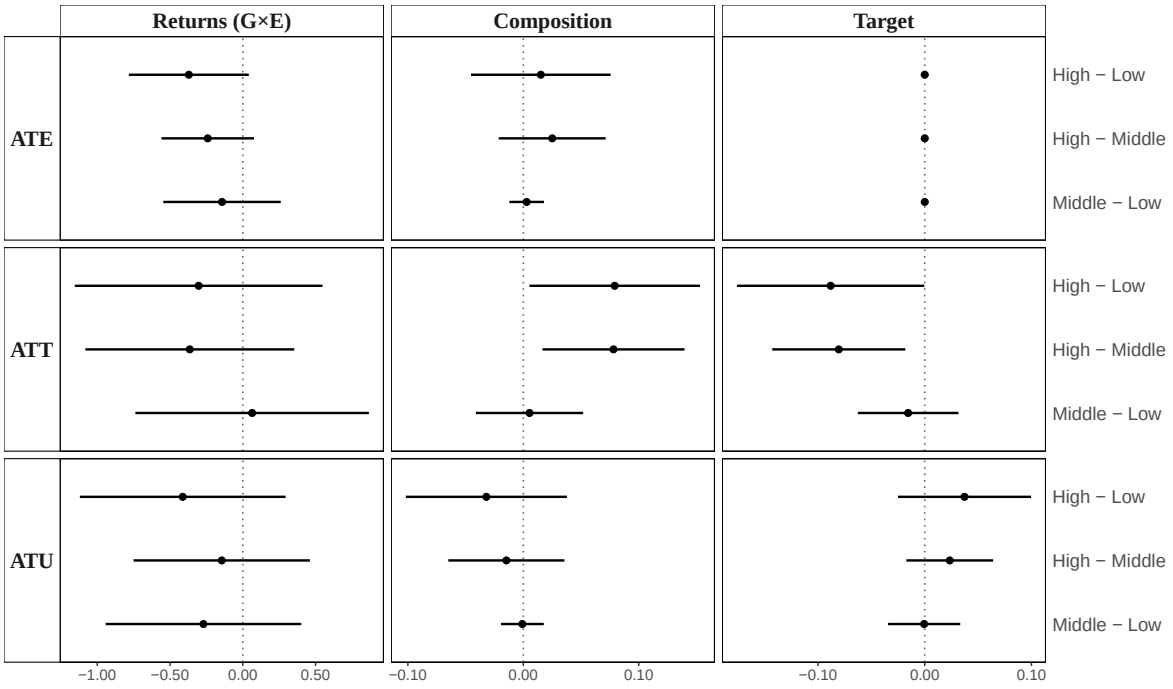


Figure 10: Decomposition of Pairwise Differences in Treatment Effects ($\tau^g - \tau^{g'}$) across EA PGI Groups. This figure reports pairwise differences in treatment effects between the High, Middle, and Low EA-PGI groups. Following Proposition 3, each difference is decomposed into returns, composition, and target channels: the returns ($G \times E$) channel captures differences in marginal returns at a common resistance margin, the composition channel differences in where each group is located along the pooled resistance scale, and the target channel differences in the estimand-specific weights placed on each resistance margin. The ATE target channel is zero by construction. Columns correspond to these three components, while rows report results for the ATE, ATT, and ATU. Effects are in log-wage points. Points denote point estimates and horizontal bars indicate 95% confidence intervals based on 1,000 bootstrap replications. Precise values are reported in Table A.15.

gene–environment correlation (rGE), on the other. The first captures heterogeneity in the causal payoff to education across genetic groups at a common resistance margin. The latter captures differences in educational sorting and in the way marginal gains are aggregated across the latent resistance distribution. This decomposition helps clarify whether observed between-group differences in returns reflect genuine heterogeneity in causal gains or instead arise because groups are located differently along the selection margin.

We apply this framework to the post-war expansion of higher education in the United Kingdom, combining UK Biobank data with a newly constructed geocoded dataset of higher-education institutions and an individual-level measure of changes in access to universities during childhood. We find a sizeable average return to a university degree together with substantial heterogeneity in returns along the resistance margin. Once group-specific MTEs are expressed on a common pooled scale, estimated returns to a university degree are larger for individuals in the low EA PGI group than for those in the high EA PGI group. The point estimates attribute most of the gap to heterogeneity in returns rather than selection-related weighting.

These findings suggest that, in the historical UK context we study, the expansion of university access had consequences beyond allowing more individuals to obtain a degree. The larger estimated wage gains among individuals with a lower genetic predisposition to education are consistent with higher education playing a compensatory role, helping to offset some of the earnings inequalities associated with individuals' starting points. Although the estimated difference remains imprecise and our comparison is limited to the available common support, the findings suggest that widening educational opportunity can also change who benefits most from it.

References

Abdellaoui, Abdel, Loic Yengo, Karin J. H. Verweij, and Peter M. Visscher. 2023. "15 Years of GWAS Discovery: Realizing the Promise." *American Journal of Human Genetics* 110 (2):179–194.

Abramitzky, Ran, Leah Platt Boustan, and Osea Giuntella. 2025. "Enclaves and Assimilation in the Age of Mass Migration: Evidence from Ethnic Catholic Churches." NBER Working Paper 33362, National Bureau of Economic Research.

Anderson, Robert. 2006. *British Universities Past and Present*. London: Hambledon Continuum.

Barcellos, Silvia H., Leandro S. Carvalho, and Patrick Turley. 2018. "Education can reduce health differences related to genetic risk of obesity." *Proceedings of the National Academy of Sciences* 115 (42):E9765–E9772.

———. 2021. "The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages, and Later-Life SES." NBER Working Paper 28750, National Bureau of Economic Research, Cambridge, MA.

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

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.

Becker, Joel, Casper A. P. Burik, Grant Goldman, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Daniel W. Belsky et al. 2021. "Resource Profile and User Guide of the Polygenic Index Repository." *Nature Human Behaviour* 5 (12):1744–1758.

Becker, Sascha O., Denni Tommasi, and Lina Zhang. 2025. "Killing Distance: How Fixed Effects Drain Power in Spatial Designs." Working paper. Available at https://sites.google.com/site/dennitommasi/.

Beloff, Michael. 1970. *The plateglass universities*. Fairleigh Dickinson University Press.

Belsky, Daniel W., Terrie E. Moffitt, David L. Corcoran, Benjamin W. Domingue, HonaLee Harrington, Sean Hogan, Renate M. Houts et al. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development." *Psychological Science* 27 (7):957–972.

Benjamin, Daniel J., David Cesarini, Patrick Turley, and Alexander Strudwick Young. 2026. "Social-Science Genomics: Progress, Challenges, and Future Directions." *Journal of Economic Literature* Forthcoming.

Berlingieri, Francesco, Christina Gathmann, and Matthias Quinckhardt. 2022. "College Openings and Local Economic Development." Discussion Paper 15364, IZA.

Biroli, Pietro, Titus Galama, Stephanie von Hinke, Hans van Kippersluis, Cornelius A. Rietveld, and Kevin Thom. 2026. "The Economics and Econometrics of Gene–Environment Interplay." *Review of Economic Studies* 93 (1):144–180.

Blundell, Richard, David Green, and Wenchao Jin. 2016. "The Puzzle of Graduate Wages." IFS Briefing Note BN185, Institute for Fiscal Studies, London.

Blundell, Richard, David A. Green, and Wenchao Jin. 2022. "The U.K. as a Technological Follower: Higher Education Expansion and the College Wage Premium." *Review of Economic Studies* 89 (1):142–180.

Bolton, Paul. 2012. "Education: Historical Statistics." Standard Note SN/SG/4252, House of Commons Library.

Braccioli, Federica, Paolo Ghinetti, Simone Moriconi, Costanza Naguib, and Michele Pellizzari. 2024. "Education Expansion, College Choice and Labour Market Success." BSE Working Paper 1419, Barcelona School of Economics.

Bratti, Massimiliano, Elena Cottini, and Paolo Ghinetti. 2022. "Education, Health and Health-Related Behaviors: Evidence from Higher Education Expansion." IZA Discussion Paper 15035, IZA Institute of Labor Economics.

Brinch, Christian N., Magne Mogstad, and Matthew Wiswall. 2017. "Beyond LATE with a Discrete Instrument." *Journal of Political Economy* 125 (4):985–1039.

Calhoun, Craig. 2014. "The Robbins Report and British Higher Education Past and Future." In *Shaping Higher Education: 50 Years After Robbins*, edited by Nicholas Barr. London: London School of Economics and Political Science, 65–86.

Card, David. 1995. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." In *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, edited by L. N. Christofides, E. K. Grant, and R. Swidinsky. University of Toronto Press, 201–222.

———. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, vol. 3, edited by Orley C. Ashenfelter and David Card, chap. 30. Elsevier, 1801–1863.

———. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (5):1127–1160.

Carneiro, Pedro, James J. Heckman, and Edward J. Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101 (6):2754–2781.

Carneiro, Pedro and Sokbae Lee. 2009. "Estimating Distributions of Potential Outcomes Using Local Instrumental Variables with an Application to Changes in College Enrollment and Wage Inequality." *Journal of Econometrics* 149 (2):191–208.

Carpentier, Vincent. 2006. "Funding in Higher Education and Economic Growth in France and the United Kingdom, 1921–2003." *Higher Education Management and Policy* 18 (3):1–26.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2016. "From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions." *Labour Economics* 41:47–60.

———. 2018. "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance." *Journal of Political Economy* 126 (6):2356–2409.

Currie, Janet and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *Quarterly Journal of Economics* 118 (4):1495–1532.

Fan, Jianqing and Irene Gijbels. 1996. *Local Polynomial Modelling and Its Applications*, *Monographs on Statistics and Applied Probability*, vol. 66. New York: Chapman & Hall.

Harden, Kathryn Paige. 2021. *The Genetic Lottery: Why DNA Matters for Social Equality*. Princeton, NJ: Princeton University Press.

Heckman, James J., H. Ichimura, and P. E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64 (4):605–654.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics* 88 (3):389–432.

Heckman, James J. and Edward J. Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences* 96 (8):4730–4734.

———. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73 (3):669–738.

Hillman, Nicholas. 2013. "From Grants for All to Loans for All: Undergraduate Finance from the Implementation of the Anderson Report (1962) to the Implementation of the Browne Report (2012)." *Contemporary British History* 27 (3):249–270.

Hollenbach, Johannes, Hendrik Schmitz, and Matthias Westphal. 2026. "Gene-environment interactions with essential heterogeneity." *Economic Journal* Advance online publication.

Ichino, Andrea, Aldo Rustichini, and Giulio Zanella. 2026. "College, Cognitive Ability, and Socioeconomic Disadvantage: Policy Lessons from the U.K. in 1960–2004." *Review of Economic Studies* Advance online publication.

Kamhöfer, Daniel A., Hendrik Schmitz, and Matthias Westphal. 2019. "Heterogeneity in Marginal Non-Monetary Returns to Higher Education." *Journal of the European Economic Association* 17 (1):205–244.

Kay, Ronald. 1985. *UCCA: Its Origins and Development 1950–85*. Cheltenham: Universities Central Council on Admissions.

Kling, Jeffrey R. 2001. "Interpreting Instrumental Variables Estimates of the Returns to Schooling." *Journal of Business & Economic Statistics* 19 (3):358–364.

Kowalski, Amanda E. 2023. "Reconciling Seemingly Contradictory Results from the Oregon Health Insurance Experiment and the Massachusetts Health Reform." *Review of Economics and Statistics* 105 (3):646–664.

Kweon, Hyeokmoon, Casper A. P. Burik, Richard Karlsson Linnér, Ronald de Vlaming, Aysu Okbay, Daphne Martschenko, Kathryn Paige Harden, Thomas A. DiPrete, and Philipp D. Koellinger. 2020. "Genetic Fortune: Winning or Losing Education, Income, and Health." Tinbergen Institute Discussion Paper 2020-053/V, Tinbergen Institute.

Layard, Richard. 2014. "What Was the World Like Then? The Context in 1963." In *Shaping Higher Education: 50 Years After Robbins*, edited by Nicholas Barr. London: London School of Economics and Political Science, 13–22.

Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet 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.

Mann, Horace. 1848. *Twelfth Annual Report of the Secretary of the Massachusetts Board of Education*. Boston, MA: Dutton and Wentworth, State Printers. Report for 1848; printed in 1849.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. New York: Columbia University Press for the National Bureau of Economic Research.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky. 2018. "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters." *Econometrica* 86 (5):1589–1619.

Mogstad, Magne and Alexander Torgovitsky. 2018. "Identification and Extrapolation of Causal Effects with Instrumental Variables." *Annual Review of Economics* 10:577–613.

———. 2024. "Instrumental Variables with Unobserved Heterogeneity in Treatment Effects." In *Handbook of Labor Economics*, vol. 5. Elsevier, 1–114.

Moser, Claus. 2014. "The Report." In *Shaping Higher Education: 50 Years After Robbins*, edited by Nicholas Barr. London: London School of Economics and Political Science, 23–32.

Mostafavi, Hakhamanesh, Arbel Harpak, Ipsita Agarwal, Dalton Conley, Jonathan K. Pritchard, and Molly Przeworski. 2020. "Variable Prediction Accuracy of Polygenic Scores within an Ancestry Group." *eLife* 9:e48376.

Mountjoy, Jack. 2022. "Community Colleges and Upward Mobility." *American Economic Review* 112 (8):2580–2630.

———. 2026. "Marginal Returns to Public Universities." *Quarterly Journal of Economics* 141 (1):429–497.

Nybom, Martin. 2017. "The Distribution of Lifetime Earnings Returns to College." *Journal of Labor Economics* 35 (4):903–952.

Okbay, Aysu, Yeda Wu, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Seyed Moeen Nehzati, Julia Sidorenko 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.

Osikominu, Aderonke, Gregor Pfeifer, and Tim Ruberg. 2026. "Characterizing the Returns to STEM: Marginal and Policy-Relevant Treatment Effects." Discussion Paper 160/26, ROCKWOOL Foundation Berlin.

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

Pratt, John. 1997. *The Polytechnic Experiment: 1965–1992*. Buckingham: Society for Research into Higher Education and Open University Press.

Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich. 2006. "Principal components analysis corrects for stratification in genome-wide association studies." *Nature Genetics* 38 (8):904–909.

Psacharopoulos, George. 1994. "Returns to investment in education: A global update." *World Development* 22 (9):1325–1343.

Renshaw, Edward F. 1960. "Estimating the Returns to Education." *Review of Economics and Statistics* 42 (3):318–324.

Robbins, Lionel. 1963. *Higher Education: Report of the Committee Appointed by the Prime Minister Under the Chairmanship of Lord Robbins, 1961-63*. London: Her Majesty's Stationery Office (HMSO). Cmnd. 2154.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2):135–146.

Shattock, Michael. 2012. *Making Policy in British Higher Education 1945–2011*. Maidenhead: McGraw-Hill/Open University Press.

Shorrocks, Anthony F. 2013. "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value." *Journal of Economic Inequality* 11 (1):99–126.

Southall, H. R. and D. Dorling. 2021. "Great Britain Historical Database: Census Data: Education Statistics, 1951–1961." UK Data Service, Study Number 4552. Data collection, 2nd edition.

Sudlow, Cathie, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey et al. 2015. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLoS Medicine* 12 (3):e1001779.

Suhonen, Tuomo and Hannu Karhunen. 2019. "The Intergenerational Effects of Parental Higher Education: Evidence from Changes in University Accessibility." *Journal of Public Economics* 176:195–217.

van Alten, S., B. W. Domingue, J. Faul, T. Galama, and A. T. Marees. 2024. "Reweighting UK Biobank corrects for pervasive selection bias due to volunteering." *International Journal of Epidemiology* 53 (3):dyae054.

Vilhjálmsson, Bjarni J., Jian Yang, Hilary K. Finucane, Alexander Gusev, Sara Lindström, Stephan Ripke, Giulio Genovese et al. 2015. "Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores." *American Journal of Human Genetics* 97 (4):576–592.

von Hinke, Stephanie and Nicolai Vitt. 2024. "An analysis of the accuracy of retrospective birth location recall using sibling data." *Nature Communications* 15 (1):2665.

Westphal, Matthias, Daniel A. Kamhöfer, and Hendrik Schmitz. 2022. "Marginal College Wage Premiums Under Selection Into Employment." *Economic Journal* 132 (646):2231–2272.

Willetts, David. 2014. "Introduction." In *Shaping Higher Education: 50 Years After Robbins*, edited by Nicholas Barr. London: London School of Economics and Political Science, 1–12.

Zhou, Xiang. 2020. *localIV: Estimation of Marginal Treatment Effects using Local Instrumental Variables*. R package version 0.3.1.

Zimmerman, Seth D. 2014. "The Returns to College Admission for Academically Marginal Students." *Journal of Labor Economics* 32 (4):711–754.

A Proof of Propositions

Proof of Proposition P-1

Let F and $\{F^g\}_{g\in\mathcal{G}}$ be continuous and strictly increasing on a common support $\mathcal{S}\subset\mathbb{R}$ with densities $f>0$ and $f^g>0$ on the interior of $\mathcal{S}$, and mixture $F=\sum_g \pi_g F^g$ where $\pi_g=\Pr(G=g)$, $\sum_g \pi_g=1$. For $u\in(0,1)$ define

$$u^g(u) \;\equiv\; F^g\big(F^{-1}(u)\big), \qquad \omega_g(u) \;\equiv\; (u^g)'(u).$$

We prove items (i)–(v).

(i) Monotonicity and derivative. Since F^{-1} and F^g are strictly increasing, the composition $u\mapsto u^g(u)=F^g\big(F^{-1}(u)\big)$ is strictly increasing. By the chain rule and the inverse–function theorem,

$$(u^g)'(u) \;=\; f^g\big(F^{-1}(u)\big)\cdot\big(F^{-1}\big)'(u) \;=\; \frac{f^g\big(F^{-1}(u)\big)}{f\big(F^{-1}(u)\big)} \;\equiv\; \omega_g(u) \;>\; 0,$$

because $f, f^g>0$ on the interior of $\mathcal{S}$.

(ii) Endpoints. Under common support and no boundary atoms, $F^{-1}(0^+)=\inf\mathcal{S}$ and $F^{-1}(1^-)=\sup\mathcal{S}$. Continuity of F^g on $\mathcal{S}$ implies

$$\lim_{u\downarrow 0} u^g(u)=F^g\big(F^{-1}(0^+)\big)=0, \qquad \lim_{u\uparrow 1} u^g(u)=F^g\big(F^{-1}(1^-)\big)=1,$$

hence $u^g(0)=0$ and $u^g(1)=1$.

(iii) Mixture (barycentric) identity. Let $u\in(0,1)$ and set $s=F^{-1}(u)$. Using the mixture relation $F=\sum_g \pi_g F^g$,

$$\sum_g \pi_g\, u^g(u) \;=\; \sum_g \pi_g\, F^g\big(F^{-1}(u)\big) \;=\; \sum_g \pi_g F^g(s) \;=\; F(s) \;=\; u.$$

Differentiating both sides w.r.t. u and using item (i) yields $\sum_g \pi_g (u^g)'(u)=1$ for all $u\in(0,1)$ where derivatives exist.

(iv) Posterior composition. Let $U\equiv F(S)$ so that $U\sim\mathrm{Unif}(0,1)$. For $u\in(0,1)$ and $s=F^{-1}(u)$,

$$\Pr(G=g\mid U=u)=\frac{\pi_g\, f^g(s)}{f(s)}=\pi_g\,\frac{f^g\big(F^{-1}(u)\big)}{f\big(F^{-1}(u)\big)}=\pi_g\,(u^g)'(u)=\pi_g\,\omega_g(u),$$

by Bayes' rule and item (i).

(v) Calibration identity. Let $P(Z) \equiv \Pr(E = 1 \mid Z) = F(\zeta(Z))$ and $P^g(Z) \equiv \Pr(E = 1 \mid Z, G = g) = F^g(\zeta(Z))$. For any Z,

$$u^g\big(P(Z)\big) = F^g\Big(F^{-1}\big(F(\zeta(Z))\big)\Big) = F^g\big(\zeta(Z)\big) = P^g(Z),$$

where we use that $F^{-1}(F(\cdot))$ is the identity on the support of $\zeta(Z)$ under continuity/overlap. Hence $u^g \circ P = P^g$ a.s., i.e., u^g calibrates the group propensity to the pooled propensity.

Proof of Proposition P-2

Fix a group $g \in \mathcal{G}$. Let $S \equiv \psi(V)$ denote the selection index with pooled CDF F and group CDF F^g, both continuous and strictly increasing on a common support. Define the pooled and group ranks

$$U \equiv F(S) \in [0,1], \qquad U^g \equiv F^g(S) \in [0,1].$$

By the probability integral transform, $U \sim \text{Unif}(0,1)$ and, on $\{G = g\}$, $U^g \sim \text{Unif}(0,1)$. The group–specific MTE is

$$MTE_g(v) \;\equiv\; \mathbb{E}[Y(1) - Y(0) \mid U^g = v,\; G = g]\,, \qquad v \in [0,1],$$

and the generic group functional is

$$\tau^g \;=\; \int_0^1 MTE_g(v)\, \omega_\tau^g(v)\, dv,$$

where $\omega_\tau^g(\cdot)$ is the (nonnegative) target weight for the parameter of interest (ATE, TT, TUT, 2SLS, etc.).

Define the *warp* and *tilt* functions on $[0,1]$:

$$u^g(u) \;\equiv\; F^g\big(F^{-1}(u)\big), \qquad \omega_g(u) \;\equiv\; (u^g)'(u) \;=\; \frac{f^g(F^{-1}(u))}{f(F^{-1}(u))}.$$

By continuity and strict monotonicity of F and F^g on a common support, $u^g : [0,1] \to [0,1]$ is strictly increasing with $u^g(0) = 0$, $u^g(1) = 1$, and $\omega_g(u) > 0$ almost everywhere.

Change of variables. Start from the definition of τ^g and substitute $v = u^g(u)$. Since u^g is strictly increasing with the stated endpoints, the substitution is valid and $dv = (u^g)'(u)\, du = \omega_g(u)\, du$. Hence

$$\tau^g = \int_0^1 MTE_g(v)\, \omega_\tau^g(v)\, dv = \int_0^1 MTE_g\big(u^g(u)\big)\, \omega_\tau^g\big(u^g(u)\big)\, \omega_g(u)\, du.$$

This is exactly the pooled–rank representation claimed in Proposition P-2:

$$\tau^g = \int_0^1 MTE_g\big(u^g(u)\big)\; \omega_\tau^g\big(u^g(u)\big)\; \omega_g(u)\; du.$$

Proof of Proposition P-3

For each group g, write the group-specific treatment functional on the pooled rank domain as

$$\tau^g \;=\; \int_0^1 A_g(u)\, B_g(u)\, C_g(u)\, du, \qquad A_g(u) \equiv MTE_g\big(u^g(u)\big),\;\; B_g(u) \equiv \omega_\tau^g\big(u^g(u)\big),\;\; C_g(u) \equiv \omega_g(u),$$

and let $W_g(u) \equiv B_g(u)C_g(u)$. The difference of interests is

$$\tau^g - \tau^{g'} = \int_0^1 \Big(A_g W_g - A_{g'} W_{g'}\Big)\, du.$$

Step 1 (Two–way symmetric decomposition for a product). For any real numbers (x_0, y_0) and (x_1, y_1),

$$x_1 y_1 - x_0 y_0 = \frac{y_1 + y_0}{2}\,(x_1 - x_0) + \frac{x_1 + x_0}{2}\,(y_1 - y_0). \tag{11}$$

Proof of (11). Add and subtract $\frac{1}{2}x_1y_0 + \frac{1}{2}x_0y_1$:

$$\begin{aligned} x_1y_1 - x_0y_0 &= \tfrac{1}{2}(x_1y_1 - x_1y_0) + \tfrac{1}{2}(x_1y_1 - x_0y_1) + \tfrac{1}{2}(x_1y_0 - x_0y_0) + \tfrac{1}{2}(x_0y_1 - x_0y_0) \\ &= \tfrac{1}{2}(y_1 - y_0)x_1 + \tfrac{1}{2}(x_1 - x_0)y_1 + \tfrac{1}{2}(x_1 - x_0)y_0 + \tfrac{1}{2}(y_1 - y_0)x_0 \\ &= \tfrac{y_1+y_0}{2}(x_1 - x_0) + \tfrac{x_1+x_0}{2}(y_1 - y_0). \end{aligned}$$

Step 2 (Outer two–way step on (A, W)). Apply (11) pointwise in u with $x_1 = A_g(u)$, $x_0 = A_{g'}(u)$ and $y_1 = W_g(u)$, $y_0 = W_{g'}(u)$:

$$A_g W_g - A_{g'} W_{g'} = \frac{W_g + W_{g'}}{2}\,(A_g - A_{g'}) + \frac{A_g + A_{g'}}{2}\,(W_g - W_{g'}).$$

Integrating over $u \in [0, 1]$ gives

$$\tau^g - \tau^{g'} = \int_0^1 \frac{W_g + W_{g'}}{2}\,(A_g - A_{g'})\, du + \int_0^1 \frac{A_g + A_{g'}}{2}\,(W_g - W_{g'})\, du. \tag{12}$$

The first integral in (12) is exactly the *MTE channel* in the proposition.

Step 3 (Inner two–way step on (B, C) inside W). Again by (11), pointwise in u with $x_1 = B_g$, $x_0 = B_{g'}$ and $y_1 = C_g$, $y_0 = C_{g'}$,

$$W_g - W_{g'} \;=\; B_gC_g - B_{g'}C_{g'} = \frac{C_g + C_{g'}}{2}\,(B_g - B_{g'}) + \frac{B_g + B_{g'}}{2}\,(C_g - C_{g'}).$$

Substitute this identity into the second integral in (12):

$$\int_0^1 \frac{A_g + A_{g'}}{2} (W_g - W_{g'})\, du = \int_0^1 \frac{A_g + A_{g'}}{2} \cdot \frac{C_g + C_{g'}}{2} (B_g - B_{g'})\, du + \int_0^1 \frac{A_g + A_{g'}}{2} \cdot \frac{B_g + B_{g'}}{2} (C_g - C_{g'})\, du.$$

The first term on the right-hand side is the *Target-Weight channel* and the second is the *Composition (tilt) channel*.

Conclusion. Combining Step 2 and Step 3 yields the claimed exact decomposition:

$$\begin{aligned} \tau^g - \tau^{g'} \;=\; & \int_0^1 \left(\frac{W_g + W_{g'}}{2} \right) (A_g - A_{g'})\, du \\ & + \int_0^1 \left(\frac{A_g + A_{g'}}{2} \cdot \frac{C_g + C_{g'}}{2} \right) (B_g - B_{g'})\, du \\ & + \int_0^1 \left(\frac{A_g + A_{g'}}{2} \cdot \frac{B_g + B_{g'}}{2} \right) (C_g - C_{g'})\, du, \end{aligned}$$

with $A_g = MTE_g\big(u^g(u)\big)$, $B_g = \omega_\tau^g\big(u^g(u)\big)$, $C_g = \omega_g(u)$, and $W_g = B_g C_g$. Each channel is linear in its own difference and uses simple arithmetic averages of the other factors; no cross-terms remain. This completes the proof.

B Detailed Description of the Data

B.1 Imputation of Log Hourly Wages

Since the UK Biobank does not contain direct information on individual income, we proxy income using an imputed occupational wage measure following Kweon et al. (2020, Online Appendix I). Their procedure combines respondents' 4-digit occupation codes (SOC) and working-hours information in UKB with external wage data from the UK Labour Force Survey (LFS) and the Annual Survey of Hours and Earnings (ASHE), estimating a parametric model of log wages and then predicting wages in UKB. The validity of this approach is assessed out-of-sample in the British Household Panel Survey (BHPS), where imputed and observed wages show reasonably strong correlations. The appendix also reports strong internal consistency between the two imputed wage measures within the UKB.[43]

B.2 Higher Education Institutions (HEI) Dataset

We construct a novel Higher Education Institution (HEI) dataset that records the timing, location, and characteristics of colleges and universities in the United Kingdom at both the institution and campus level. For each entry, we record the institution name, campus name, whether the campus is the main campus, latitude and longitude, institution type, and source information. We also collect detailed temporal information, including the institution's founding year, the year in which teaching activity began at a given campus where relevant, the year in which university status was granted, and the year of closure where applicable. In addition, the dataset contains information on student numbers in 2020, public versus private status, and curriculum. Curriculum is recorded both as free text and through a categorical classification that groups campuses into broad types such as generalist institutions, core-subject institutions, specialised institutions, theological institutions, hospitals, and campuses with no undergraduate teaching. The dataset and accompanying documentation will be made publicly available upon publication.

The dataset was assembled by combining official and institutional sources. Information on legal recognition, degree-awarding powers, and university status was checked against official UK government guidance on recognised degree-awarding institutions (GOV·UK), which directs users to the relevant official registers and lists for each UK nation.[44] Historical information on founding dates, predecessor institutions, campus openings, and institutional development was collected primarily from institutions' official history pages and related archival material hosted on university websites. Campus addresses were taken from official institution websites and, where needed, converted into geographic coordinates using mapping tools for geocoding. Since many institutions operate multiple campuses, we identify the main campus whenever possible, while retaining additional campuses when they provide relevant degree programmes. This approach ensures that our measure of proximity captures meaningful access to higher education rather than administrative definitions of institutions. These choices are important for

[43] See Online Appendix I of Kweon et al. (2020) for full details on its construction and validation.

[44] For England, official information on registration, degree-awarding powers, and university title is maintained by the Office for Students (OfS) in the OfS Register.

measurement, as they affect the mapping between geographic access and actual educational opportunities. We retain the original source associated with each entry in order to ensure transparency and to facilitate future extensions and updates of the dataset.

For the empirical analyses, we use these records to reconstruct the university network available at each point in time. The relevant unit is a distinct geographic university access point, identified by institution, campus coordinates, and the year in which it enters the university network. Duplicate records representing the same access point are counted only once. A campus enters the network in its coded university-status year, which need not coincide with the institution's original founding date or the date teaching began at that location. It is considered active in a given year if it has entered the university network by that year and has not subsequently closed. The Open University is excluded. This campus-based definition underlies the calculation of distance to the nearest university and the construction of the reduction-in-distance instrument used throughout the paper.

We also distinguish different sources of expansion within the university network. The *Pure New* category comprises campuses belonging to eight purpose-built post-war universities: Sussex, East Anglia, York, Lancaster, Essex, Kent, Warwick, and Stirling. All remaining entries are classified as *Other Source*, including institutions that acquired university status through conversion and subsequent campus expansions. In the source-specific analyses, the eight colleges of advanced technology (CATs) converted to universities in the 1960s (Aston, Bath, Bradford, Brunel, City, Loughborough, Salford, and Surrey) are treated as a separate category and excluded from *Other Source*. These categories distinguish the creation of new universities from other forms of entry into the university network, which may have built on pre-existing higher-education provision. Consequently, an entry into the university network does not necessarily represent the opening of an entirely new educational institution or campus.

Table A.1: Variable Definitions and Sources

Variable	Definition	Source	Level	Notes
Treatment and Instrument				
University degree	Binary indicator for holding a college or university degree	UKB	Individual	
Reduction in distance	Distance (km) from the birthplace to the nearest university campus operating at birth minus the corresponding distance at age 15	HEI, UKB	Individual	
Genetic Measures				
EA PGI	Educational Attainment PGI, standardised to mean zero and unit variance in the analysis sample	UKB	Individual	
Genetic principal components	First 20 genetic principal components	UKB	Individual	Account for population stratification.
Outcomes				
Log hourly wage	Imputed log hourly wage based on the individual's current or most recent occupation (SOC)	UKB, ASHE, LFS	Individual	Appendix B.1.
Controls				
A. Individual Characteristics				
Birth year	Year of birth	UKB	Individual	
Age	Age of the individual at recruitment	UKB	Individual	
Male	Binary indicator for male sex	UKB	Individual	
Number of siblings	Total number of siblings	UKB	Individual	
B. Early-Life Local Context				
Universities within 10 km	Number of universities operating within 10 km of the birthplace at birth	HEI, UKB	Individual	
Urban district	Binary indicator for an urban (vs. rural) district of birth	1951 Census	District	
Educational level	Share of adult males in full-time education until at least age 15	1951 Census	District	Female tabulations only available for towns above 50,000 inhabitants.
Log population	Log population of the district of birth	GBHD	District	Fixed to 1951 values.
Log population density	Log population density of the district of birth	GBHD	District	Fixed to 1951 values.
Population growth	Population growth between 1946 and 1951 (%)	GBHD	District	
Share low SES	Share of the occupied population in Social Classes IV and V	1951 Census	District	Partly skilled and unskilled occupations.
Share high SES	Share of the occupied population in Social Classes I and II	1951 Census	District	Professional, managerial, and related occupations.
Unemployment rate	Unemployment rate in the district of birth (1951)	Labour Gazette	District	
Weights				
UKB inverse-probability weight	Inverse-probability weight based on UK Census microdata	UKB	Individual	Constructed by van Alten et al. (2024); used in weighted descriptive and IV analyses. MTE estimation is unweighted.

Notes: This table provides definitions and data sources for the variables used in the paper. The EA PGI is described in Section 3.3; wage imputation is described in Appendix B.1. UKB = UK Biobank. HEI = Higher Education Institutions Dataset (Appendix B.2). GBHD = Great Britain Historical Database. ASHE = Annual Survey of Hours and Earnings. LFS = Labour Force Survey. PGI = polygenic index.

Table A.2: Summary Statistics by Graduate Status

Variable	(1) Overall	(2) Non-graduates	(3) Graduates	(4) Difference (3) − (2)
Outcome				
Log hourly wage	2.654	2.559	2.821	0.262***
	(0.345)	(0.335)	(0.293)	(0.003)
Treatment, Genetic Predisposition, and Instrument				
University degree	0.361	0.000	1.000	1.000
	(0.480)	(0.000)	(0.000)	
EA PGI	-0.018	-0.197	0.297	0.494***
	(0.986)	(0.950)	(0.969)	(0.009)
Reduction in distance (km)	3.848	3.492	4.477	0.985***
	(11.316)	(10.605)	(12.447)	(0.177)
*Controls (**X**)*				
A. Individual Characteristics				
Birth year	1959.058	1959.143	1958.908	-0.235***
	(6.307)	(6.278)	(6.355)	(0.064)
Age	49.264	49.187	49.401	0.215***
	(6.299)	(6.269)	(6.349)	(0.052)
Age squared	2466.660	2458.648	2480.812	22.165***
	(638.705)	(636.034)	(643.152)	(5.175)
Male	0.521	0.524	0.516	-0.008**
	(0.500)	(0.499)	(0.500)	(0.004)
Number of siblings	1.891	1.988	1.719	-0.270***
	(1.481)	(1.566)	(1.300)	(0.013)
B. Early-Life Local Context				
Universities within 10 km before birth	2.317	2.290	2.366	0.076
	(6.305)	(6.179)	(6.521)	(0.160)
Urban district (1951)	0.876	0.879	0.871	-0.008
	(0.329)	(0.326)	(0.335)	(0.005)
Educational level in district of birth	0.240	0.234	0.251	0.017***
	(0.080)	(0.078)	(0.083)	(0.002)
Log population (1951)	4.318	4.385	4.200	-0.185***
	(1.359)	(1.368)	(1.336)	(0.044)
Log population density (1951)	7.488	7.532	7.411	-0.120***
	(1.341)	(1.319)	(1.375)	(0.024)
Population growth (1946–1951)	0.025	0.024	0.025	0.001***
	(0.020)	(0.019)	(0.021)	(0.000)
Share low SES	0.296	0.300	0.290	-0.011***
	(0.073)	(0.071)	(0.074)	(0.002)
Share high SES	0.172	0.166	0.181	0.015***
	(0.068)	(0.066)	(0.072)	(0.001)
UR (1951)	1.064	1.076	1.044	-0.032
	(0.851)	(0.862)	(0.832)	(0.022)
District UR source	0.684	0.701	0.653	-0.048***
	(0.465)	(0.458)	(0.476)	(0.008)
Observations	157109	94878	62231	

Notes: Entries are means, with standard deviations in parentheses. The final column reports the mean among university graduates minus the mean among non-graduates, with district-of-birth-clustered robust standard errors. Means, standard deviations and contrasts use UKB weights. The University contrast is mechanical, so no significance stars are reported for it. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.3: Distribution of the Reduction in Distance to the Nearest University

	(1) Overall	(2) Positive Reduction	(3) Negative Reduction	(4) No Reduction
Mean reduction	3.848	13.275	−0.537	0
Standard deviation	11.316	17.758	0.723	0
Minimum	−3.948	0.001	−3.948	0
25th percentile	0	1.281	−0.452	0
Median	0	4.288	−0.317	0
75th percentile	1.010	19.391	−0.159	0
Maximum	96.897	96.897	−0.001	0
Observations	157,109	48,087	1,834	107,188
Share of sample	100.00%	29.06%	1.70%	69.24%

Notes: The table reports summary statistics for the reduction in distance to the nearest university in the estimation sample of this paper. The reduction is calculated as the difference, in kilometres, between the distance to the nearest university active at birth and the distance to the nearest university active by age 15. Means, standard deviations, percentiles, and shares of the sample use the UKB inverse-probability weights; observation counts and extrema are unweighted.

C More on the University Expansion

In the decade following the Second World War, British higher education remained small, selective, and geographically concentrated (Anderson, 2006; Carpentier, 2006). University provision was centred on Oxford and Cambridge, London, the ancient Scottish universities, the civic "redbrick" universities,[45] and a relatively small number of university colleges. In 1962, only around 4 percent of each generation entered university, with access varying sharply by social background, sex, region, and prior schooling. For many prospective students, attending university therefore meant leaving home and bearing substantial accommodation, travel, and foregone-earnings costs. Geographic proximity nevertheless remained economically relevant in this initially sparse system. Large parts of the country had no nearby university, so the arrival of a new institution could materially alter the effective cost and feasible set of university options facing local students.

The postwar period brought a major expansion in the supply and geographic coverage of university education. Its landmark policy statement was the 1963 Robbins Report (Robbins, 1963). However, the expansion did not begin with Robbins. During the 1950s, existing universities expanded and the sector began to extend beyond its pre-war geography, while increasing numbers of school-leavers acquired the qualifications required for university entry. By the end of the decade, the number of qualified school-leavers was growing faster than the supply of places, and over the five years preceding the Robbins inquiry the ratio of first-year university entrants to qualified school-leavers fell by roughly one quarter (Layard, 2014, pp. 16–17). The approaching post-war baby-boom cohorts added further pressure. The University Grants Committee (UGC) was therefore already planning substantial expansion before the Report was published and, by the beginning of the 1960s, envisaged around 170,000 university places by the early 1970s (Willetts, 2014, p. 8). The expansion should consequently be understood as a gradual post-war process that gathered momentum during the 1950s and accelerated sharply during the following decade.

An important component of this early expansion was the creation of new universities. Existing institutions were reluctant to accommodate the entire projected increase in student numbers, leading the UGC in the late 1950s to support a new generation of universities. Sussex opened in 1961, followed by East Anglia and York in 1963, Lancaster and Essex in 1964, and Kent and Warwick in 1965. These "plate-glass" universities represented an important departure from the traditional development of British higher education, in which locally established colleges often expanded gradually before acquiring university status (Beloff, 1970). Several were instead founded directly as universities on newly developed campuses (Shattock, 2012, p. 43); (Willetts, 2014, p. 9). Because the university network was initially sparse, the earliest additions could substantially improve access in areas located far from an existing university. As the network expanded, subsequent additions took place against an increasingly dense and geographically dispersed system.

Where this new capacity was located was itself an explicit policy concern. In selecting the first wave of new universities, the UGC gave considerable weight to cities without existing university provision, while the availability of suitable sites, staff housing, and student accommodation also entered into decisions

[45] Including institutions such as Birmingham, Leeds, Liverpool, Manchester, and Sheffield.

over prospective locations. The Robbins Committee subsequently argued that future expansion should place greater weight on large cities and major centres of population, both to serve larger student catchment areas and to exploit their proximity to cultural institutions, industry, commerce, medicine, government, and research facilities (Robbins, 1963, pp. 162–163). Local political support and the presence of existing higher-education institutions also affected which places entered the university sector. The geography of expansion therefore reflected national objectives, local circumstances, and the pre-existing distribution of higher education rather than a uniform increase in capacity across the country.

The expansion coincided with reforms that made other dimensions of university access increasingly national. Following the Anderson Report, the Education Act 1962 introduced a substantially more standardised system of tuition and maintenance support for eligible students, replacing sizeable differences in the arrangements operated by local authorities (Hillman, 2013; Willetts, 2014). Admissions were reorganised at around the same time: the Universities Central Council on Admissions (UCCA), established in 1961, became operational for the 1964 entry cycle (Kay, 1985; Willetts, 2014). These reforms reduced some of the financial and administrative differences facing prospective students across the country, but physical access remained uneven. In 1961–62, around 20 percent of full-time university students still lived at home, while a further 52 percent relied on private lodgings (Robbins, 1963, pp. 194–195). Travel, relocation, and accommodation therefore remained important features of university attendance even as the financial and administrative framework for entry became increasingly national.

Published in October 1963, the Robbins Report provided the central rationale for accelerating and broadening the expansion. Rather than determining the size of higher education from forecasts of demand for particular types of graduate, the Committee argued that provision should respond to qualified demand. Its central principle—described in the Report as an "axiom"—was that "courses of higher education should be available for all those who are qualified by ability and attainment to pursue them and who wish to do so" (Robbins, 1963, p. 8). The Committee concluded that existing capacity fell substantially short of the number of young people able and willing to pursue higher education and recommended rapid further expansion. The government accepted its principal recommendations immediately after publication, giving political force to and extending a process already under way (Willetts, 2014; Moser, 2014).

Expansion after Robbins was not confined to newly founded universities. Established higher-education institutions also entered the university sector through changes in status. Most notably, the Colleges of Advanced Technology (CATs) and related institutions—including Aston, Bath, Bradford, Brunel, City, Loughborough, Salford, and Surrey—acquired university status during the 1960s (Calhoun, 2014). These conversions built on pre-existing local technical capacity and therefore differed from the arrival of the newly founded "plate-glass" universities, but they nevertheless expanded the formal university sector and the set of institutions offering university-level degrees. More generally, the post-war expansion combined genuinely new universities with the incorporation of existing institutions and the continued growth of established universities. Appendix Figure A.1 illustrates this evolution in our reconstructed university network, distinguishing new university campuses from other entries into the sector and showing the resulting increase in the number of active university locations over time. The scale of the transformation

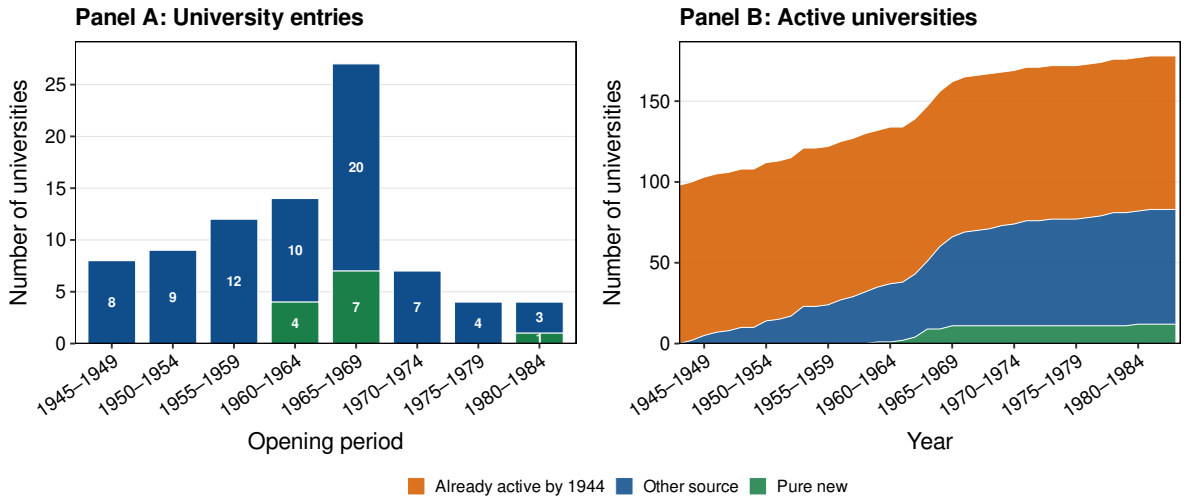


Figure A.1: University expansion by source and entry period. Panel A reports the number of distinct university campuses entering the analytic university network in each five-year period from 1945–1949 to 1980–1984. Panel B reports the corresponding stock of active university campuses in each year from 1945 through 1985. A campus is considered active when its coded university-status year is no later than the relevant calendar year and its closure year is no earlier than that year. Campuses are classified as Already Active by 1944 when they had entered the university sector by 1944. Pure New comprises campuses belonging to the eight pre-specified purpose-built postwar universities: the Universities of Sussex, East Anglia, York, Lancaster, Essex, Kent, Warwick, and Stirling. Other Source comprises all remaining coded entries into the university network, including university-status conversions and subsequent campus expansions. The unit counted is a distinct geographic university access point, identified by institution, coordinates, and coded university-status year; duplicate records representing the same access point are counted once. This campus-based definition corresponds to the geographic network used to calculate distance to the nearest university and construct the reduction-in-distance instrument. Entry years indicate when a campus enters the coded university network and do not necessarily coincide with the physical foundation date of the campus. The Open University is excluded.

is also evident in aggregate degree production: the annual number of first degrees awarded rose from 22,426 in 1960 to 51,189 in 1970 (Bolton, 2012; Calhoun, 2014). The cohorts studied in this paper therefore came of age during a period in which both the scale and the geography of British university provision were changing markedly across places and over time.

Table A.4: Predetermined District Characteristics and the Geography of University Expansion

	(1)	(2)	(3)	(4)
	Any opening	Any opening	Opening count	Opening count
Baseline universities within 10 km	0.028**	0.028*	0.065**	0.079*
	(0.012)	(0.014)	(0.031)	(0.041)
Urban district	-0.007	-0.010	0.033	0.029
	(0.021)	(0.021)	(0.036)	(0.035)
Log population, 1951	0.035***	0.034***	0.068***	0.072***
	(0.009)	(0.010)	(0.020)	(0.022)
Log population density, 1951	0.004	0.006	-0.016	-0.007
	(0.009)	(0.009)	(0.023)	(0.018)
Population growth, 1946–1951	-0.006	-0.006	-0.008	-0.006
	(0.004)	(0.004)	(0.007)	(0.007)
Unemployment rate, 1951	-0.009***	-0.010	-0.011**	-0.025*
	(0.003)	(0.008)	(0.005)	(0.015)
Share educated over age 15	0.015**	0.013**	0.020*	0.019
	(0.006)	(0.007)	(0.010)	(0.012)
Low-SES share	0.008	0.007	0.009	0.010
	(0.006)	(0.006)	(0.009)	(0.009)
High-SES share	0.005	0.006	0.008	0.013
	(0.006)	(0.006)	(0.011)	(0.010)
Region fixed effects	No	Yes	No	Yes
Num. districts	1,468	1,468	1,468	1,468
R^2	0.086	0.087	0.107	0.114

Notes: The unit of observation is a 1951 England–Wales local-government district. The analysis uses the full district universe, subject only to complete data for the district characteristics included in the regressions. University events are campus-level openings or university-status acquisitions from 1945 to 1980, defined using the same university-status date field as the instrument construction and assigned to districts using campus coordinates and 1951 district boundaries. Baseline university provision is the number of universities active in 1944 within 10 km of an interior point of each district (a point guaranteed to lie within its 1951 boundary). All predictors except the urban indicator, including baseline university provision, are standardised using their means and standard deviations across the full district universe, so coefficients give the change associated with a one-standard-deviation increase. The 1951 unemployment rate uses the district rate where available and the fixed regional rate otherwise, so unemployment does not reduce the sample. District characteristics are measured at or near the beginning of the expansion period, so the table provides descriptive evidence on observable placement patterns rather than a test of instrument exogeneity. Columns (1)–(2) are linear probability models for any opening; columns (3)–(4) are OLS regressions of the number of openings. Heteroskedasticity-robust standard errors are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

D Instrument Validity

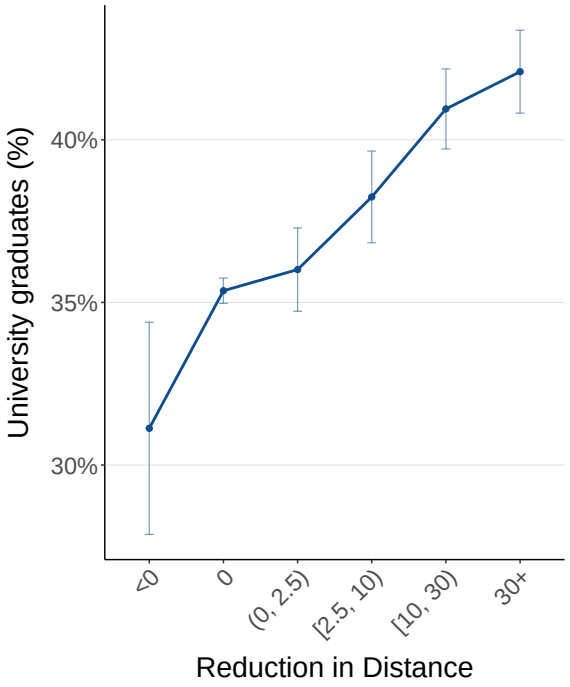


Figure A.2: Degree Attainment by Bins of the Reduction in Distance to University. This figure plots the weighted proportion holding a university degree (in percent) across bins of the reduction in distance to university. Bins are defined to distinguish (i) negative reductions in distance, (ii) no change in distance, and (iii) positive reductions partitioned into sufficiently large and broadly comparable groups. The horizontal axis reports the resulting exposure bins, while the vertical axis reports the corresponding weighted proportion holding a university degree. The plotted values correspond to the "Mean University" column of Table A.5, expressed as percentages. Vertical bars report 95 percent confidence intervals based on the weighted effective sample size in each bin. Means are weighted using UK Biobank inverse-probability weights.

Table A.5: Descriptive Monotonicity Check

$\Delta Dist$ bin	(1) Mean $\Delta Dist$	(2) Mean University	(3) Residualised University	(4) Δ	(5) Δ_0	(6) Observations
$[-10, 0)$	−0.47	0.311	−0.028	—	−0.025	1,595
0	0.00	0.354	−0.003	0.025	—	117,223
$(0, 2.5)$	1.01	0.360	−0.002	0.001	0.001	9,685
$[2.5, 10)$	5.01	0.382	−0.001	0.001	0.002	8,670
$[10, 30)$	18.16	0.410	0.023	0.024	0.026	11,057
$[30, \infty)$	47.56	0.421	0.035	0.013	0.038	9,993

Notes: This table provides a descriptive check of the monotonicity condition for the reduction in distance to university ($\Delta Dist$). University completion is first residualised with respect to the full set of baseline controls, genetic principal components, and birth-region-by-cohort fixed effects, excluding $\Delta Dist$ itself, using UK Biobank inverse-probability weights. Observations are grouped into bins designed to distinguish (i) negative reductions in distance, (ii) no change in distance, and (iii) positive reductions partitioned into sufficiently large and broadly comparable groups. For each bin, the table reports the weighted mean of the instrument, the weighted proportion holding a university degree, the weighted mean of the residualised degree indicator, and the number of observations. Δ reports the change in mean residualised university completion relative to the preceding exposure bin, while Δ_0 reports the difference relative to the zero-exposure bin ($\Delta Dist = 0$).

Table A.6: First-Stage Robustness to Alternative Geographic Specifications

	(1) Region × cohort FE	(2) Additive FE	(3) Additive FE + region trends	(4) County × cohort FE
$\Delta Dist$	0.014***	0.011***	0.012***	0.010***
	(0.002)	(0.002)	(0.002)	(0.003)
Observations	157,109	157,109	157,109	157,109
R^2	0.090	0.087	0.088	0.108
First-stage F	35.47	28.89	31.68	10.78
Controls (X)	Yes	Yes	Yes	Yes

Notes: The dependent variable is an indicator for holding a university degree. $\Delta Dist$ denotes the reduction in distance to the nearest university, measured in 10 km units. Column (1) presents the baseline specification with birth-region-by-cohort fixed effects. Column (2) replaces these interactions with additive birth-region and cohort fixed effects. Column (3) additionally allows for birth-region-specific linear trends across cohorts. Column (4) replaces the baseline fixed effects with birth-county-by-cohort fixed effects. Standard errors clustered by numerical district of birth are reported in parentheses, and inference uses district-cluster degrees of freedom. The first-stage F-statistic is the district-cluster-robust Wald statistic for the excluded instrument. All columns use the same sample, baseline controls, 20 genetic principal components, and UK Biobank inverse-probability weights. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.7: Placebo tests using post-entry reductions in distance

	(1)	(2)	(3)
	25 → 30	25 → 35	25 → 40
Panel A: First stage			
Dependent variable: University degree			
$\Delta Dist$	0.003	0.004	0.004
	(0.004)	(0.004)	(0.005)
Placebo F-statistic	0.51	0.74	0.69
Joint Wald test (p-value)		0.843	
Panel B: Reduced form			
Dependent variable: Log wage			
$\Delta Dist$	0.001	0.002	0.001
	(0.003)	(0.002)	(0.002)
Joint Wald test (p-value)		0.831	
Observations		157,109	

Notes: Each placebo reduction is measured per 10 km relative to the nearest university available at birth. Only universities gaining university status strictly after age 25 and by the indicated upper age are admitted; all intermediate openings outside that window are excluded. All specifications include the updated baseline controls, EA PGI and its interaction with sex, birth-region-by-cohort fixed effects, all 20 exported genetic principal-component controls, and UK Biobank inverse-probability weights. Standard errors clustered by numerical birth-district ID are reported in parentheses. Inference uses birth-district cluster degrees of freedom. Panel A reports placebo regressions with degree attainment as the outcome; Panel B reports placebo reduced-form regressions for log average wage. The joint Wald-test row reports the p-value for the null that the coefficients on all three placebo windows are jointly equal to zero, accounting for the overlap across windows. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.8: Balance of Predetermined Characteristics with Respect to University Access

	(1)	(2)	(3)
	EA PGI	Male	Number of siblings
$\Delta Dist$	0.007	-0.001	-0.008
	(0.005)	(0.002)	(0.005)
Mean dep. var.	0.022	0.476	1.860
Observations	157,109	157,109	157,109

Notes: Each column reports a separate weighted OLS regression of the listed predetermined outcome on the reduction in distance. $\Delta Dist$ is measured per 10 km. Standard errors clustered by district of birth are reported in parentheses. All specifications include the baseline controls and birth-region-by-cohort fixed effects. When the tested outcome is itself included among the baseline controls, it is omitted from the control set in that specification. The mean dependent variable is the unweighted mean in the corresponding regression sample. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.9: First-Stage and Reduced-Form Robustness to Alternative Instrument Definitions

	Alternative Exposure Windows				
	(1)	(2)	(3)	(4)	(5)
	Birth→11	Birth→12	Birth→13	Birth→14	Birth→15
Panel A: First stage					
Dependent variable: University degree					
Δ*Dist*	0.014***	0.014***	0.014***	0.014***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
First-stage *F*	38.37	37.06	35.58	35.93	35.47
Panel B: Reduced form					
Dependent variable: Log wage					
Δ*Dist*	0.006***	0.006***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	157,109	157,109	157,109	157,109	157,109

	Alternative University Definitions		
	(6)	(7)	(8)
	All institutions	Pure new	Other source
Panel A: First stage			
Dependent variable: University degree			
Δ*Dist*	0.014***	0.020***	0.012***
	(0.002)	(0.004)	(0.003)
First-stage *F*	35.47	22.11	18.20
Panel B: Reduced form			
Dependent variable: Log wage			
Δ*Dist*	0.006***	0.009***	0.004***
	(0.001)	(0.002)	(0.001)
Observations	157,109	157,109	157,109

Notes: Each column reports a separate specification using the indicated definition of the university-access instrument. Δ*Dist* denotes the reduction in distance to the nearest university and is measured in 10 km units. The first-stage dependent variable is an indicator for holding a university degree, while the reduced-form dependent variable is log average wage. The exposure-window specifications vary the age by which university openings are allowed to affect the instrument. The university-definition specifications split the university entries behind the baseline birth-to-age-15 instrument by source, following the classification in Appendix B.2: *Pure new* uses only the campuses of the eight purpose-built post-war universities (Sussex, East Anglia, York, Lancaster, Essex, Kent, Warwick, and Stirling), and *Other source* uses all remaining entries, including conversions to university status and campus expansions, except the eight colleges of advanced technology (CATs) converted to universities in the 1960s (Aston, Bath, Bradford, Brunel, City, Loughborough, Salford, and Surrey), which form a separate category that is not reported. Each source-specific instrument is the nonnegative reduction in distance generated by entries of that source between birth and age 15, relative to the distance at birth to the full university network. The source-specific instruments are estimated separately and are not additive components of the baseline instrument: for each individual, the largest of the pure-new, other-source, and CAT instruments equals the positive part of the baseline instrument, which can be negative because of closures. All specifications use the same estimation sample, baseline controls, birth-region-by-cohort fixed effects, 20 genetic principal components, and UK Biobank inverse-probability weights. Standard errors clustered by numerical district of birth are reported in parentheses. The first-stage *F*-statistic is the district-cluster-robust Wald statistic for the excluded instrument. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.10: First-Stage and Reduced-Form Robustness to Alternative Functional Forms

	(1) Baseline	(2) Quadratic	(3) Log
Panel A: First stage			
Dependent variable: University degree			
$\Delta Dist$	0.014***	0.013**	
	(0.002)	(0.007)	
$\Delta Dist^2$		0.000	
		(0.001)	
$\Delta \log(Dist)$			0.014**
			(0.006)
First-stage F	35.47	21.49	5.45
Panel B: Reduced form			
Dependent variable: Log wage			
$\Delta Dist$	0.006***	0.008**	
	(0.001)	(0.004)	
$\Delta Dist^2$		-0.000	
		(0.001)	
$\Delta \log(Dist)$			0.009***
			(0.003)
Observations	157,109	157,109	157,109

Notes: Each column reports a separate specification using the indicated functional form of university access. In columns (1) and (2), $\Delta Dist$ denotes the reduction in distance to the nearest university and is measured in 10 km units. Column (2) additionally includes its squared term. Column (3) uses the change in log distance to the nearest university. All specifications use the same estimation sample, baseline controls, birth-region-by-cohort fixed effects, 20 genetic principal components, and UK Biobank inverse-probability weights. Standard errors clustered by numerical district of birth are reported in parentheses. The first-stage F-statistic is the district-cluster-robust Wald statistic for the excluded instrument; in column (2), it corresponds to the joint test that the coefficients on $\Delta Dist$ and $\Delta Dist^2$ are zero. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.11: First-Stage Robustness to Alternative Link Functions

	(1) LPM	(2) Probit	(3) Logit
Panel A: Baseline			
Average marginal effect (Z)	0.014^{***}	0.014^{***}	0.014^{***}
	(0.003)	(0.003)	(0.002)
Fitted-score overlap	[-0.05, 0.81]	[0.06, 0.80]	[0.07, 0.80]
Fitted-score overlap (0.25% per tail)	[0.13, 0.67]	[0.16, 0.68]	[0.16, 0.68]
Fitted-score overlap (0.50% per tail)	[0.16, 0.64]	[0.18, 0.65]	[0.18, 0.65]
Panel B: $Z \times X^b$ Included			
Average marginal effect (Z)	0.015^{***}	0.015^{***}	0.015^{***}
	(0.002)	(0.002)	(0.002)
Fitted-score overlap	[-0.05, 0.82]	[0.06, 0.80]	[0.07, 0.80]
Fitted-score overlap (0.25% per tail)	[0.13, 0.67]	[0.16, 0.68]	[0.16, 0.68]
Fitted-score overlap (0.50% per tail)	[0.16, 0.64]	[0.18, 0.65]	[0.18, 0.65]
Observations	157,109	157,109	157,109
Region × cohort FE	Yes	Yes	Yes
Controls (X)	Yes	Yes	Yes
Genetic PCs	20	20	20

Notes: Reduction in distance is measured per 10 km. All models use the same sample, baseline controls, birth-region-by-cohort fixed effects, and UK Biobank inverse-probability weights. For the LPM, the reported average marginal effect is the coefficient on the reduction in distance; for the Probit and Logit models, entries report average marginal effects. Delta-method standard errors, clustered by numerical birth-district ID, are reported in parentheses. Inference uses birth-district cluster degrees of freedom. Fitted-score overlap is the intersection of the fitted-score ranges for graduate and non-graduate participants. Trimming is the percentage of observations removed from each tail within each treatment group, as in Table A.12. Panel B interacts the instrument with the pre-birth local education share, high-SES share, number of siblings, 1951 unemployment, and urban status. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.12: Empirical Common Support of the Propensity Score

	(1) Trimming (% per tail)	(2) CS_{min}	(3) CS_{max}
Panel A: Overlap between Treated and Untreated			
Pooled: $P(Z)$	0	0.081	0.782
	0.10	0.161	0.713
	0.25	0.190	0.681
	0.50	0.211	0.659
Low EA PGI: $P^1(Z)$	0	0.055	0.635
	0.10	0.087	0.550
	0.25	0.103	0.515
	0.50	0.120	0.494
Middle EA PGI: $P^2(Z)$	0	0.099	0.755
	0.10	0.160	0.698
	0.25	0.189	0.670
	0.50	0.208	0.640
High EA PGI: $P^3(Z)$	0	0.184	0.853
	0.10	0.270	0.802
	0.25	0.298	0.781
	0.50	0.321	0.766
Panel B: Common Support on the Pooled Resistance Scale			
Pooled-score overlap across tertiles	0	0.104	0.765
	0.10	0.171	0.704
	0.25	0.197	0.671
	0.50	0.216	0.653
Common pooled u after group mappings	0	0.111	0.758
	0.10	0.172	0.697
	0.25	0.202	0.667
	0.50	0.222	0.646

Notes: CS denotes an empirical common-support interval. Panel A reports the intersection of the fitted-score supports of university graduates ($E = 1$) and non-graduates ($E = 0$), overall for the pooled propensity score and separately by EA PGI tertile for the group-specific scores. All Panel B bounds are expressed on the pooled resistance scale. Its first block intersects treated/untreated pooled-score overlap across all three tertiles. Its second block additionally requires each mapped coordinate $\hat{u}^g(u)$ to lie within its group's native treated/untreated overlap; endpoints are reported on the estimator's 100-point pooled grid. Trimming is the percentage removed from each tail within each treatment-status population whose supports are intersected: the baseline MTE specification uses 0.50% per tail (1.00% total). The fitted scores and warp mappings are held fixed across the trimming rows. The figures additionally require availability in all 1,000 bootstrap draws, giving a displayed common interval of $[0.232, 0.636]$.

E Benchmark estimation of GxE

As a benchmark, we estimate a two-stage least squares (2SLS) model of the gene–environment interaction ($G \times E$) between university education and genetic predisposition to education on wages. We present the specification using the continuous standardised EA PGI and also estimate a version using EA PGI tertiles to facilitate comparison with the MTE analysis in the main text.[46] For the continuous specification, we estimate:

$$Y_i = \beta_0 + \beta_1 G_i + \beta_2 E_i + \beta_3 (G_i \times E_i) + \gamma' X_i + \lambda_{a(i),c(i)} + \varepsilon_i, \tag{13}$$

where Y_i denotes imputed log hourly wages, E_i indicates degree attainment, G_i is the standardised EA PGI, and X_i contains the baseline controls. The term $\lambda_{a(i),c(i)}$ denotes geography-by-cohort fixed effects, where $a(i)$ refers to the individual's area of birth. Our baseline specification uses birth-region-by-cohort fixed effects, while the more geographically demanding specifications replace these with birth-county-by-cohort fixed effects.[47] In this linear specification, β_3 captures how the estimated return to a university degree varies with genetic predisposition to education.

Because degree attainment and its interaction with the EA PGI may be endogenous, we instrument $(E_i, G_i \times E_i)$ using the reduction in distance to the nearest university, Z_i, and its interaction with G_i. The corresponding first-stage equations are:

$$E_i = \pi_0 + \pi_1 G_i + \pi_2 Z_i + \pi_3 (Z_i \times G_i) + \delta' X_i + \lambda_{a(i),c(i)} + v_{1i}, \tag{14}$$

$$(G_i \times E_i) = \theta_0 + \theta_1 G_i + \theta_2 Z_i + \theta_3 (Z_i \times G_i) + \eta' X_i + \lambda_{a(i),c(i)} + v_{2i}. \tag{15}$$

Table A.13 reports the benchmark 2SLS estimates. Under the standard IV assumptions, these coefficients provide a linear-IV benchmark for the causal return to a university degree and its heterogeneity with respect to the EA PGI. With heterogeneous treatment effects, however, the interaction coefficient need not compare treatment effects for the same complier population across values of G, which motivates the MTE framework developed in the main text. Across specifications, the estimated return to a university degree is positive and sizeable. In the continuous-G specifications that include the interaction term (columns (1) and (2)), the coefficient on the degree indicator is around 0.5–0.6, while the interaction between degree attainment and the EA PGI is negative and statistically significant (−0.093 and −0.104, respectively). Thus, within this linear IV specification, a one-standard-deviation increase in the EA PGI is associated

[46] In the tertile specification, we replace the continuous EA PGI and its interaction with degree attainment with indicators for the medium and high EA PGI groups and their interactions with degree attainment, leaving the low group as the reference category. We instrument degree attainment and its interactions with these indicators using the reduction in university distance and its corresponding interactions, retaining the same controls and fixed effects.

[47] All specifications include the baseline individual- and district-level controls used in the main analysis. Specifications including genetic measures additionally control for the first 20 genetic principal components to account for population stratification (Price et al., 2006). Following Biroli et al. (2026), we interact demeaned controls with the main regressors to allow their effects to vary.

Table A.13: Returns to a University Degree: 2SLS Estimates

	Continuous G						Discrete G
	$G \times E$		G		No G		
	(1) Region	(2) County	(3) Region	(4) County	(5) Region	(6) County	(7) Region
University degree	0.512***	0.588***	0.503***	0.576***	0.440***	0.438**	0.515***
	(0.076)	(0.115)	(0.075)	(0.114)	(0.101)	(0.189)	(0.128)
Degree × EA PGI	-0.093**	-0.104***					
	(0.041)	(0.038)					
EA PGI	0.034*	0.029	-0.001	-0.009			
	(0.017)	(0.019)	(0.008)	(0.013)			
Degree × EA PGI Medium							-0.052
							(0.124)
Degree × EA PGI High							-0.168
							(0.146)
EA PGI Medium							0.020
							(0.036)
EA PGI High							0.073
							(0.062)
Num.Obs.	157109	157109	157109	157109	157109	157109	157109
R2 Adj.	0.188	0.195	0.188	0.195	0.162	0.170	0.183
First-stage robust Wald F	19.25	5.53	35.47	10.78	25.35	8.86	12.43
Controls (X)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	No	Yes	No	Yes	No	Yes
County FE	No	Yes	No	Yes	No	Yes	No
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Area × cohort trend	No	No	No	No	No	No	No
Genetic PCs	20	20	20	20	0	0	20

Notes: The outcome is log hourly wage. University degree and its interactions with G are instrumented with the reduction in distance and its interactions with G. Discrete G denotes EA PGI tertiles, with Low as the reference category. Columns labelled Region include additive birth-region and cohort fixed effects; columns labelled County replace the region effects with county-of-birth fixed effects, so that identification relies only on variation within counties, which weakens the first stage. Columns (5)–(6) exclude the EA PGI and the genetic principal components. Standard errors, clustered by district of birth, are reported in parentheses. UK Biobank inverse-probability weights are applied. The first-stage F is the district-cluster-robust Wald statistic for the joint null that all excluded instruments are zero in the first-stage equation for university degree; in specifications with interactions, it jointly tests the reduction in distance and its interactions with G. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

with a reduction in the estimated graduate wage premium of approximately 9–10 log points. This pattern is consistent with a negative gene–environment interaction in the returns to a university degree.

The remaining specifications yield a broadly similar picture. When the EA PGI enters only additively (columns (3) and (4)), its coefficient is small and statistically insignificant, while the estimated effect of holding a university degree remains positive and sizeable. Excluding genetic controls altogether (columns (5) and (6)) similarly leaves the estimated return to a university degree positive. Finally, the discrete-EA-PGI specification in column (7) is directionally consistent with the continuous specification: the estimated interactions are negative for both the middle and high EA PGI tertiles, with the high-tertile interaction larger in magnitude. These estimates are, however, imprecise and neither interaction is statistically distinguishable from zero.

Taken together, the benchmark estimates provide suggestive evidence that the return to a university degree declines with genetic predisposition to education. However, conventional interaction-IV estimates need not compare the same complier or resistance margins across genetic groups. This motivates the MTE analysis in the main text, which allows treatment-effect heterogeneity and educational selection to vary flexibly across the resistance distribution.

F Sensitivity of the Warp and Tilt Estimates

Table A.14: Warp Sensitivity to Propensity-Score Specification

	Deviation from baseline					
	Low		Middle		High	
	Sup	RMSE	Sup	RMSE	Sup	RMSE
Propensity-score link						
Probit	–	–	–	–	–	–
Logit	0.003	0.001	0.005	0.001	0.006	0.002
LPM	0.007	0.003	0.009	0.002	0.011	0.004

Notes: This table reports sensitivity of the estimated group-specific warp functions, $\widehat{u}^g(u)$, to alternative propensity-score specifications. Deviations are computed relative to the baseline specification, defined as a Probit propensity-score model. "Sup" denotes the supremum absolute deviation on the evaluation grid relative to the baseline warp, $\sup_u \left|\widehat{u}^{g,\text{alt}}(u) - \widehat{u}^{g,\text{base}}(u)\right|$, and "RMSE" denotes the root mean squared deviation over the same grid. All deviations use the same supported pooled-u grid, $[0.242, 0.626]$, shared by all specifications. Low, Middle, and High refer to EA PGI tertiles. LPM denotes linear probability model, with fitted scores clipped to $[10^{-6}, 1 - 10^{-6}]$.

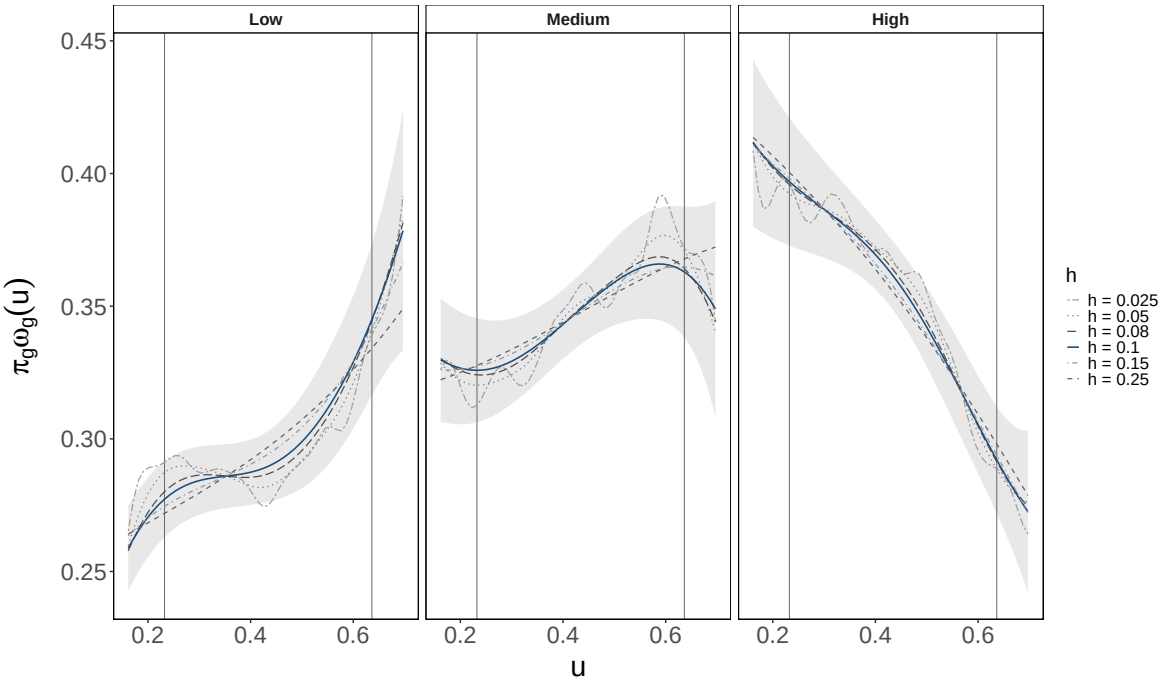


Figure A.3: Sensitivity of the Mixture Weights to the Bandwidth. The figure plots the estimated mixture weights, $\hat{\pi}_g\,\hat{\omega}_g(u)$, for each EA PGI tertile under alternative bandwidths h used to smooth the derivative of the warp, $\omega_g(u) = (u^g)'(u)$ (Proposition 1); as in Figure 7, the weights sum to one across groups at each u. The shaded region is the pointwise 95% normal confidence band, centred on the original-sample estimate and constructed from 1,000 bootstrap replications, for the baseline bandwidth ($h = 0.10$) used throughout the paper. Vertical lines indicate the common-overlap interval of u used for cross-group MTE comparisons, approximately $[0.232, 0.636]$. Values outside this interval are shown for diagnostic purposes and are not used to compute cross-group treatment-effect functionals.

G Additional MTE Results and Sensitivity Analyses

Table A.15: Decomposition of Differences in Treatment Effects

	Returns	Target	Composition	Total
Panel A: ATE				
Middle vs Low	-0.143	–	0.003	-0.140
	(0.206)		(0.008)	(0.207)
High vs Low	-0.371*	–	0.015	-0.355*
	(0.210)		(0.031)	(0.213)
High vs Middle	-0.241	–	0.025	-0.216
	(0.162)		(0.024)	(0.163)
Panel B: ATT				
Middle vs Low	0.064	-0.016	0.005	0.054
	(0.409)	(0.024)	(0.024)	(0.411)
High vs Low	-0.303	-0.088**	0.079**	-0.312
	(0.434)	(0.045)	(0.038)	(0.435)
High vs Middle	-0.364	-0.081**	0.078**	-0.366
	(0.366)	(0.032)	(0.031)	(0.366)
Panel C: ATU				
Middle vs Low	-0.270	-0.001	-0.001	-0.272
	(0.343)	(0.017)	(0.009)	(0.343)
High vs Low	-0.413	0.037	-0.032	-0.407
	(0.361)	(0.032)	(0.036)	(0.361)
High vs Middle	-0.144	0.023	-0.015	-0.136
	(0.309)	(0.021)	(0.026)	(0.308)

Notes: This table decomposes differences in treatment effects, $\tau^g - \tau^{g'}$, across EA PGI tertiles into MTE, target, and composition channels following Proposition 3. Original-sample estimates with bootstrap standard errors in parentheses from 1,000 individual-resampled replications. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, from two-sided normal tests using bootstrap standard errors. A dash indicates that the target-weight channel does not apply to the ATE (it is zero by construction): the reported returns and composition terms use the two-channel decomposition with group weights normalised over the available common interval. ATT and ATU use the three-channel decomposition. Restricted-support quadrature and normalisation mean that the ATE decomposition total can differ slightly from the difference between the group ATE entries in Table 3.

G.1 Sensitivity Tests

Table A.16: Returns to a University Degree: Sensitivity to Bandwidth Choice

		Resistance Support				Treatment Effects†		
	h	$u^g_{\min}$	$u^g_{\max}$	$u_{\min}$	$u_{\max}$	ATE	ATT	ATU
Panel A: Pooled MTE								
Unconditional	0.05			0.21	0.66	0.485*** (0.077)	0.714*** (0.174)	0.321** (0.144)
	0.10			0.21	0.66	0.495*** (0.070)	0.729*** (0.163)	0.330** (0.134)
	0.15			0.21	0.66	0.491*** (0.068)	0.718*** (0.159)	0.332** (0.132)
Panel B: Group-Conditional MTE								
Low EA PGI	0.05	0.14	0.48	0.22	0.65	0.691*** (0.187)	0.766** (0.361)	0.638** (0.289)
	0.10	0.14	0.48	0.22	0.65	0.684*** (0.178)	0.746** (0.345)	0.639** (0.280)
	0.15	0.14	0.48	0.22	0.65	0.687*** (0.175)	0.742** (0.337)	0.646** (0.281)
Middle EA PGI	0.05	0.22	0.62	0.22	0.65	0.552*** (0.124)	0.769*** (0.261)	0.408** (0.206)
	0.10	0.22	0.62	0.22	0.65	0.554*** (0.117)	0.805*** (0.246)	0.388** (0.197)
	0.15	0.22	0.62	0.22	0.65	0.543*** (0.112)	0.796*** (0.239)	0.374* (0.193)
High EA PGI	0.05	0.34	0.75	0.22	0.65	0.292** (0.128)	0.347 (0.289)	0.227 (0.256)
	0.10	0.34	0.75	0.22	0.65	0.328*** (0.123)	0.422 (0.279)	0.243 (0.244)
	0.15	0.34	0.75	0.22	0.65	0.329*** (0.120)	0.430 (0.278)	0.239 (0.239)
Panel C: Aggregation of G-Specific MTEs								
Unconditional (CS)	0.05			0.22	0.65	0.489*** (0.074)	0.715*** (0.168)	0.330** (0.137)
	0.10			0.22	0.65	0.495*** (0.070)	0.720*** (0.159)	0.339*** (0.129)
	0.15			0.22	0.65	0.492*** (0.067)	0.709*** (0.155)	0.340*** (0.127)
Aggregated groups	0.05			0.22	0.65	0.512*** (0.088)	0.627*** (0.179)	0.424*** (0.148)
	0.10			0.22	0.65	0.522*** (0.083)	0.658*** (0.171)	0.423*** (0.141)
	0.15			0.22	0.65	0.519*** (0.080)	0.656*** (0.168)	0.420*** (0.140)

Notes: This table presents the estimated treatment-effect parameters across different MTE bandwidth specifications. Panel A reports pooled unconditional treatment effects evaluated over the common support between treated and untreated individuals. Panels B and C report treatment effects evaluated over the observed common support of the resistance distribution across EA PGI groups. Bootstrap standard errors based on 1,000 replications are reported in parentheses below the point estimates. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, from two-sided normal tests using bootstrap standard errors. The same bandwidth is used in the pooled and group-specific MTE estimators within each row. Support bounds are rounded to two decimals.

†: Treatment effects are evaluated over the observed common support of the resistance distribution. These are restricted-support parameters rather than full-population effects.

Table A.17: Decomposition of Differences in Treatment Effects: Sensitivity to Bandwidth Choice

	h	Returns	Target	Composition	Total
Panel A: ATE					
		-0.412*		0.015	-0.398*
	0.05	(0.223)	–	(0.032)	(0.224)
High vs Low		-0.369*		0.015	-0.354
	0.10	(0.214)	–	(0.031)	(0.216)
		-0.371*		0.015	-0.355*
	0.15	(0.210)	–	(0.031)	(0.213)
		-0.284		0.020	-0.264
	0.05	(0.175)	–	(0.025)	(0.175)
High vs Medium		-0.251		0.024	-0.227
	0.10	(0.168)	–	(0.024)	(0.168)
		-0.241		0.025	-0.216
	0.15	(0.162)	–	(0.024)	(0.163)
		-0.136		0.002	-0.134
	0.05	(0.219)	–	(0.008)	(0.219)
Medium vs Low		-0.129		0.003	-0.127
	0.10	(0.210)	–	(0.008)	(0.211)
		-0.143		0.003	-0.140
	0.15	(0.206)	–	(0.008)	(0.207)
Panel B: ATT					
		-0.418	-0.074	0.072*	-0.419
	0.05	(0.457)	(0.049)	(0.040)	(0.459)
High vs Low		-0.316	-0.087*	0.079**	-0.324
	0.10	(0.440)	(0.046)	(0.038)	(0.441)
		-0.303	-0.088**	0.079**	-0.312
	0.15	(0.434)	(0.045)	(0.038)	(0.435)
		-0.423	-0.067*	0.068**	-0.421
	0.05	(0.387)	(0.035)	(0.033)	(0.387)
High vs Medium		-0.381	-0.080**	0.078**	-0.384
	0.10	(0.371)	(0.032)	(0.032)	(0.371)
		-0.364	-0.081**	0.078**	-0.366
	0.15	(0.366)	(0.032)	(0.031)	(0.366)
		0.011	-0.013	0.005	0.002
	0.05	(0.442)	(0.025)	(0.024)	(0.444)
Medium vs Low		0.070	-0.016	0.005	0.059
	0.10	(0.419)	(0.025)	(0.024)	(0.421)
		0.064	-0.016	0.005	0.054
	0.15	(0.409)	(0.024)	(0.024)	(0.411)
Panel C: ATU					
		-0.418	0.035	-0.029	-0.412
	0.05	(0.379)	(0.034)	(0.039)	(0.380)
High vs Low		-0.401	0.037	-0.031	-0.395
	0.10	(0.363)	(0.032)	(0.036)	(0.364)
		-0.413	0.037	-0.032	-0.407
	0.15	(0.361)	(0.032)	(0.036)	(0.361)
		-0.190	0.025	-0.017	-0.181
	0.05	(0.329)	(0.023)	(0.029)	(0.328)
High vs Medium		-0.153	0.025	-0.016	-0.145
	0.10	(0.315)	(0.021)	(0.026)	(0.314)
		-0.144	0.023	-0.015	-0.136
	0.15	(0.309)	(0.021)	(0.026)	(0.308)
		-0.227	-0.002	-0.002	-0.231
	0.05	(0.351)	(0.020)	(0.011)	(0.351)
Medium vs Low		-0.249	-0.001	-0.001	-0.251
	0.10	(0.343)	(0.017)	(0.010)	(0.343)
		-0.270	-0.001	-0.001	-0.272
	0.15	(0.343)	(0.017)	(0.009)	(0.343)

Notes: This table decomposes differences in treatment effects, $\tau^g - \tau^{g'}$, across EA PGI tertiles into MTE, target, and composition channels following Proposition 3. Entries are original-sample estimates. Bootstrap standard errors are in parentheses. All reported specifications use 1,000 bootstrap replications. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, from two-sided normal tests using bootstrap standard errors. A dash indicates that the target-weight channel does not apply to the ATE.

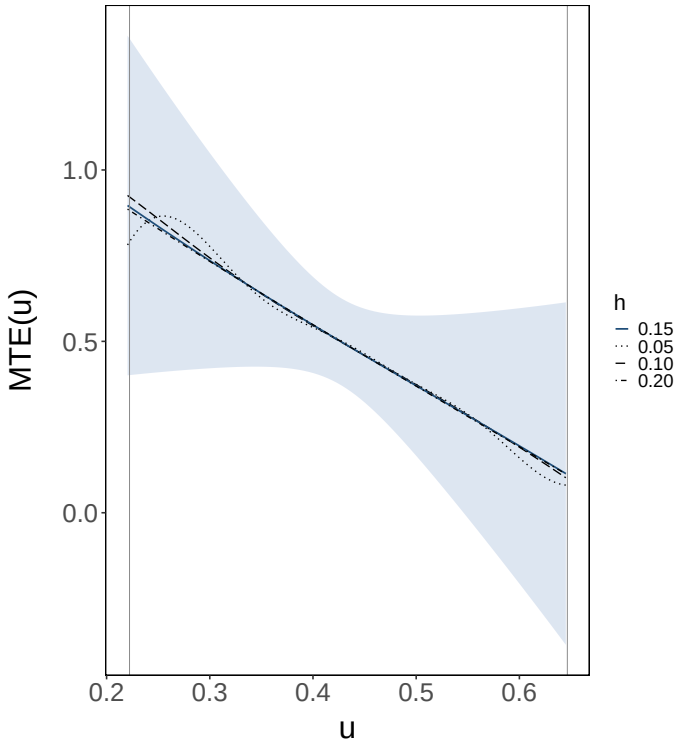


Figure A.4: Robustness of the Pooled Unconditional MTE to Bandwidth Choice. The figure reports the estimated unconditional MTE curve under alternative bandwidths used in the local polynomial smoothing step. Vertical lines mark the original-sample common support used for cross-group comparisons (unlike Figures 8 and 9, which show the interval supported in all bootstrap draws). The baseline specification is highlighted with a solid line, while the remaining curves show robustness to alternative bandwidth choices. Shaded areas denote 95% confidence intervals from 1,000 bootstrap replications for the main specification only ($h = 0.15$).

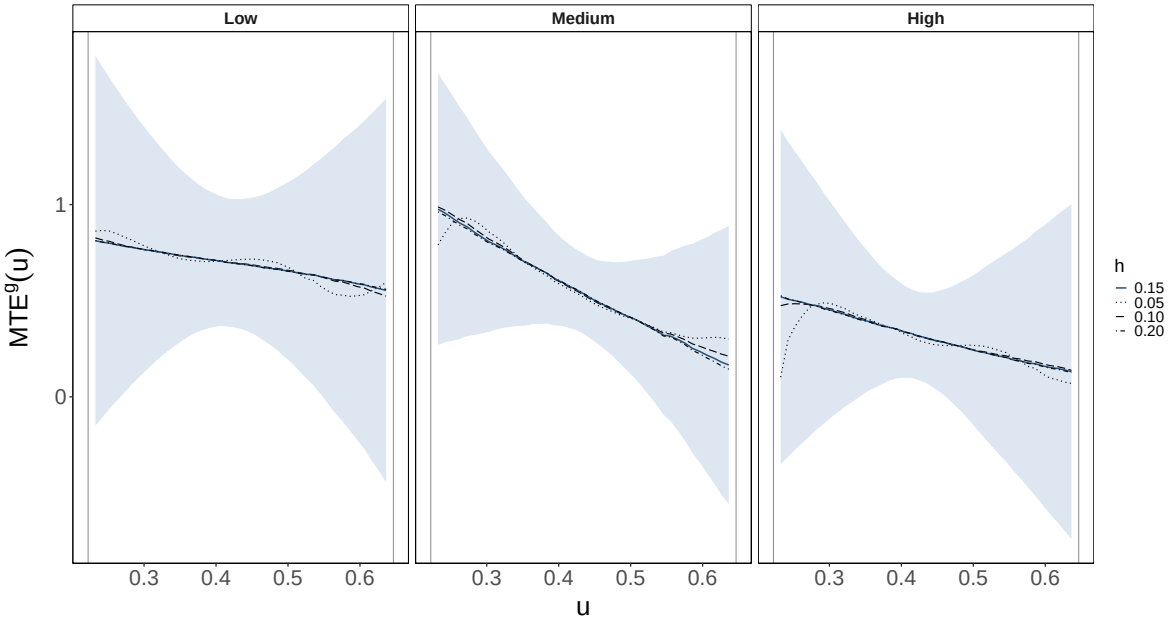


Figure A.5: Robustness of Group-Specific MTE Curves to Bandwidth Choice. The figure reports the estimated marginal treatment effect curves separately by EA PGI group under alternative bandwidths used in the local polynomial smoothing step. Panels correspond to the low, medium, and high EA PGI groups. Vertical lines mark the original-sample common support used for cross-group comparisons, as in Figure A.4. The baseline specification ($h = 0.15$) is highlighted with a solid line, while the remaining curves show robustness to alternative bandwidth choices. Shaded areas denote 95% confidence intervals, based on 1,000 bootstrap replications, for the baseline specification only.

G.2 Including continuous G in X

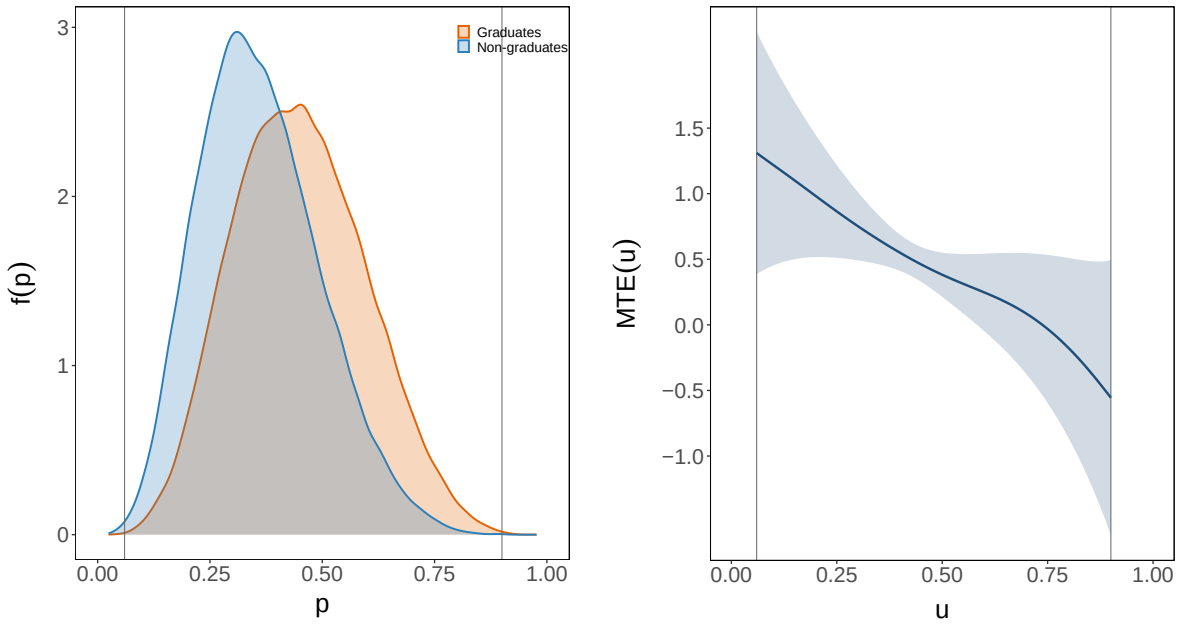


Figure A.6: Propensity Score and Pooled MTE with Continuous G. The continuous EA PGI enters both the selection and outcome equations. The left panel shows the estimated propensity-score densities for graduates and non-graduates; vertical lines mark the fixed fitting and reporting interval $[0.06, 0.90]$ rather than a trimmed overlap interval. The right panel shows the original-sample pooled MTE with pointwise 95% normal intervals from 1,000 bootstrap replications. Relative to the baseline specification described in Section 2.6, the covariate step uses a LOESS span of 0.04 and the derivative step a bandwidth of 0.10.