Who Can Have It All?

The Unequal Cost of Career & Inter-generational Mobility

Patrizia Massner Arash Nekoei*

March 27, 2026

Abstract

The standard concept of intergenerational mobility tracks movement up the occupational or income ladders but ignores the costs incurred along the way. An example of such costs arises from the career-family tradeoff. We argue that unequal career costs—meaning a steeper tradeoff for individuals from disadvantaged backgrounds—lead income mobility to overstate welfare mobility: upward mobility in professional terms translates into smaller welfare mobility as it comes at the expense of personal life. We first develop the theoretical model to evaluate the merits of this argument and its underlying hypothesis. We then empirically test the key hypothesis using Swedish data. We find that individuals from low-income families sacrifice more of their family life to achieve higher lifetime income, delaying childbearing, having fewer children, and spending less time on publicly provided parental leave per child. These unequal costs reflect a more arduous path to success, shaped by time and spatial constraints, which result in a trial-and-error life trajectory rather than guided success. Disadvantaged achievers study longer in less remunerative fields, although they are more cognitively and non-cognitively skilled at age 18, further delaying family formation, and tend to live farther from their birthplace and parents, increasing the cost of parenting.

*Patrizia: Johns Hopkins University, patrizia.massner@jhu.edu, Arash: IIES, CEPR, IZA, and RFBerlin, arash.nekoei@iies.su.se. We thank Konrad Burchardi, Raj Chetty, François Gerard, Claudia Goldin, Tatiana Homonoff, Yuhta Ishii, Markus Jäntti, Wojciech Kopczuk, Enrico Moretti, Matt Notowidigdo, Torsten Persson, Juan Carlos Suárez Serrato, and seminar participants at the Nordic Public Finance Conference 2023, NBER Spring Public Meeting 2025, Helsinki GSE Colloquium, Oslo Public-Labor Workshop, LMU, and Oxford for their comments. Valerio Passantino provided excellent research assistance.

Introduction

I heard somebody say one time: You can make it, if you try.
And some of us, we tried so hard.
Syl Johnson, 1969

In the social sciences broadly—and in economics in particular—we have extensively studied individuals who rise from the bottom to the top of the social ladder. We have often celebrated such ascents as evidence of inclusion, dynamism, and equality of opportunity. We have paid far less attention to how hard people must try, and to the costs they bear, to climb the ladder.

In this context, intergenerational mobility has emerged as a primary statistic of a society's fluidity and inclusivity in academic and policy debates.[1] The modern concept of IGM originated in sociology and describes movements across generations between social classes, typically measured using occupations.[2] In economics, the focus has shifted to examining the correlation between the income of two consecutive generations.[3]

Whether in sociology or economics, most studies of mobility carry an implicit or explicit normative interpretation. Societies with more mobility are seen as more equitable—offering equal opportunities to all—and more efficient, because they discover and harness talent regardless of background (Bell et al., 2019; Jones, 2022).

Yet a focus on mobility in outcomes—such as occupation, income, or wealth—neglects the costs associated with upward mobility. Time and effort devoted to career advancement and climbing the income ladder can crowd out other dimensions of life, such as family, leisure, and health, and thereby strain the balance between personal and professional life. A central element of this tension is the widely recognized career–family trade-off (Goldin, 2021). An ideal concept of mobility would extend beyond professional outcomes to capture overall welfare, offering a more holistic measure of how lives change across generations.

This paper tackles this challenge in three parts. We start with a theoretical model of intergenerational mobility in Part 1. The main goal of the theory is to understand the relation between income and welfare mobility. Agents derive utility from both their

[1] See Harrington (1997); Solon (1999); Beaud et al. (2013); Chetty et al. (2014); Putnam (2016); Major and Machin (2018); Mogstad and Torsvik (2023) and references therein.

[2] Starting from the ground-breaking work of Sorokin (1927), this line of research has evolved into a large field; see Nico (2021) for a recent survey. Occupation is also the primary focus of historical studies of mobility, see Van Leeuwen and Maas (2010); Long and Ferrie (2013); Berger et al. (2023).

[3] Wealth and education mobility have been studied, but on a much smaller scale. For review of the literature in economics, see Björklund and Jäntti (1997); Solon (1999); Black and Devereux (2011); Mogstad and Torsvik (2023).

professional and personal lives but face limited resources (e.g., time or energy), leading to a trade-off.

If the cost of career is the same for all, income and welfare mobility coincide. But when career costs vary by background, the two measures diverge. In fact, income mobility can be higher in a society with lower welfare mobility. Our main theoretical result is that income mobility overestimates welfare mobility if the cost of a career decreases by background. If those who climb the income ladder have lower welfare than those born and remaining at the top, they raise income mobility more than welfare mobility.

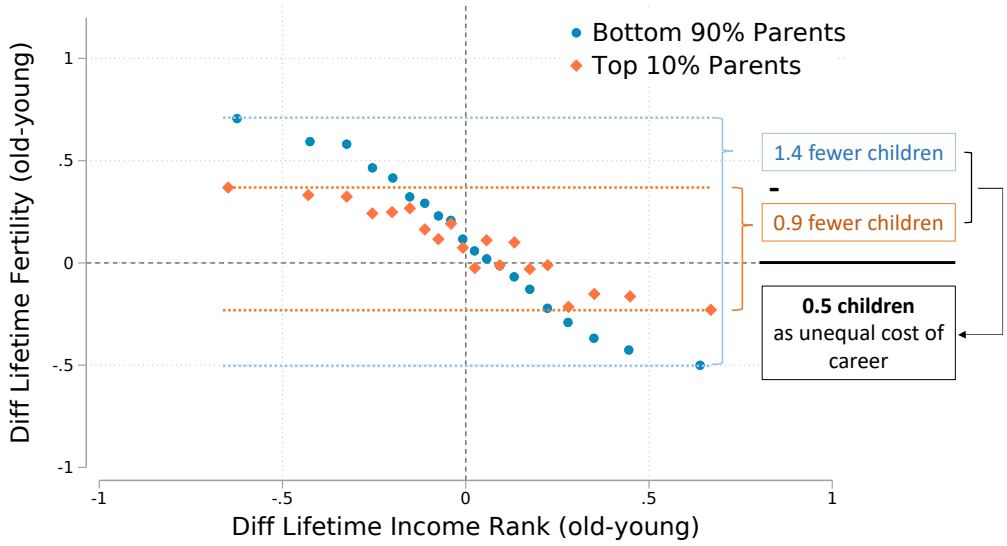
Part 2 empirically investigates the unequal cost of a career using Swedish register data. As a proxy for personal life, we focus primarily on family outcomes. Our main measures of family outcomes are the completed (lifetime) fertility rate and the time spent on parental leave per child. We complement this analysis with other aspects of personal life, such as geographical mobility. Professional life and parental background are measured through an individual's and their parents' lifetime income, similar to the recent literature on mobility (e.g., Chetty et al., 2014).

Figure 1 presents an illustrative example linking our theory to the data. Between two Swedish sisters, the one with a higher lifetime income tends, on average, to have fewer children. More importantly, this relationship varies by parental background. Among sisters whose parents are at the bottom 90% of the income distribution, moving from the bottom to the top of the income distribution results, on average, in having 1.4 fewer child. In contrast, for sisters with parents in the top decile of income, the difference is approximately 0.9 children. The difference-in-differences is half a child, which we interpret as the unequal cost of career.

Building on this example, we empirically identify the unequal cost of career by comparing the differences in family outcomes of individuals with and without careers by parental background. This approach is similar to a difference-in-differences design, where career serves as the treatment and background as the time variable, albeit without a causal interpretation. More precisely, we estimate the cross-derivative of family outcomes with respect to career and background, by regressing outcomes on two-sided fixed effects for an individual's own and parental income percentiles, along with their interactions. These interactions capture the differential family costs of a career by parental background.

The cost of career decreases with parental background. To reach the top decile of lifetime income, those with low-income parents sacrifice more of their family life: They are less likely to become parents. If they have children, they have fewer children. Self-made achievers not only have fewer children but also spend less time with each child.

Figure 1: Income vs. Fertility of Sisters by Parental Background



Note: This figure compares two sisters' lifetime fertility and income by background. It plots the difference in their completed fertility—measured as the total number of biological children born before age 45—against the difference in their lifetime income rank, based on total labor income between ages 25 and 45 within birth cohorts. The sample includes 78,462 sister pairs. "Top 10%" (or "Bottom 90%") parents indicate whether parental income—defined as the sum of the father's and mother's income—falls in the top decile or the bottom nine deciles.

We demonstrate this using the length of parental leave per child for female self-made achievers.

Part 3 shows that the unequal cost of career stems from contrasting trajectories to success, shaped by drawbacks in time and space. We show this using our diff-in-diff approach and detailed data on educational attainment, skills measured at age 18 from the military draft for all men, labor market outcomes, and geographic location.

Starting with the time dimension, we find that, to reach the top income decile, self-mades postpone childbirth—a delay substantial enough to fully account for the difference in the number of children along the intensive margin. This postponement sets off a chain of compounding disadvantages. On average, self-mades attain one additional year of education compared to peers from high-income families and enter the labor market later, as indicated by a higher age at first job.

In addition to time drawbacks, they also have spatial drawbacks. Self-mades are more likely to leave their birthplace and end up living farther from their hometowns and parents compared to privileged achievers. This suggests that disadvantaged individuals that pursue a career might get less help from grandparents with their kids, which elevates the cost of parenthood compared with achievers from high-income families.

These drawbacks in time and space stem from different educational and professional paths to success between individuals from high-income vs. low-income backgrounds. Disadvantaged achievers possess higher levels of both cognitive and noncognitive skills before entering university at age 18. This evidence suggests a lower return on education and skills for disadvantaged achievers. The lower return to years of schooling can be partially explained by the fact that they pursue less financially rewarding fields of study, as evidenced by the average income of students in these fields.

Finishing their education later, disadvantaged achievers enter the labor market later in life. Moreover, they begin their careers at lower-paying firms, indicated by low AKM firm fixed effects (Abowd et al., 1999; Card et al., 2013). In an effort to catch up, they ascend the job ladder by frequently changing jobs and enhancing the quality of firms they work for, as evidenced by increasing AKM firm fixed effects. As direct evidence of impediment, self-made achievers are half as likely to work at the same company as their parents. Working their way through the education system and labor market with more trial and error.

Prior literature We contribute to several strands of research. A long-standing literature studies the correlation of outcomes across generations, that is, intergenerational mobility. In sociology and history, occupation is the main outcome of interest (surveyed in Van Leeuwen and Maas, 2010), whereas the economics literature has focused primarily on income and less extend wealth and education (surveyed in, Solon, 1999; Jäntti et al., 2006; Björklund et al., 2009; Black and Devereux, 2011; Mogstad and Torsvik, 2023). Recent work highlights labor market characteristics, such as firms and occupations, as important pathways to success (Corak and Piraino, 2011; Boar and Lashkari, 2021; Dobbin and Zohar, 2023; Engzell and Wilmers, 2021). We build on this large body of research by looking at welfare and by studying these outcomes together with the family-life costs associated with attaining them.

Our paper also relates to the literature on the career-family trade-off (e.g., Goldin, 1997, 2014, 2021; Adda et al., 2017). Across many countries, women's earnings decline after the arrival of children (e.g., Angelov et al., 2016; Kleven et al., 2019a,b). Firm and job characteristics—such as schedule flexibility, remote work options, and family benefits—are central determinants of these penalties (Goldin, 2014; Harrington and Kahn, 2023; Casarico and Lattanzio, 2023; Hotz et al., 2018). We extend this literature by examining the extent to which the relationship between family outcomes and income varies by socioeconomic background.

The remainder of the paper unfolds as follows. Section 1 compares welfare mobility

and income mobility theoretically. Section 2 derives an empirical test of the theory, and Section 3 implements this test empirically by providing evidence of unequal costs of a career. Section 4 examines the mechanisms underlying unequal career costs. Sections 5 and 6 discuss the findings and conclude. Additional background material and auxiliary analyses are provided in the Online Appendix.

1 Theory

This section studies the theoretical link between income and welfare mobility. It conveys the core message as simply as possible, relegating the technical assumptions, details, proofs, and extensions to the Appendix A.

1.1 Setting

Individual θ cares about preferences over professional life z and personal life k. They are given lifetime endowment of L resources, $r(\theta) \in \mathbb{R}_+^L$ where $L \in \mathbb{N}$, such as time, energy, effort, or mental capacity. Resources can be used to produce z and k:

$$z = F_z(x_z, \theta), \qquad k = F_k(x_k, \theta),$$

where $x_z, x_k \in \mathbb{R}_+^L$ denote the inputs to professional and personal life, respectively. The L-resource constraint can be written as $x_z + x_k \leq T(\theta)$.[4] This constraint generates a trade-off between personal and professional life.

Assume the usual properties for preferences and production functions so that the individual's lifetime problem,

$$v(\theta) := \max_{x,y \in \mathbb{R}_+^L} u(F_z(x,\theta), F_k(y,\theta)) \quad \text{s.t.} \quad x + y \leq T(\theta). \tag{1}$$

has a unique solution.[5] Let $z(\theta)$ and $k(\theta)$ denote the resulting optimal professional and personal outcomes.[6]

Assume one-parent, one-child dynasties. We index generations by t. For example, θ_t, z_t, v_t denote, respectively, the type, the optimal professional outcome, and lifetime utility of generation t, so $z_t = z(\theta_t), v_t = v(\theta_t)$. For notational simplicity and wherever it

[4] Throughout the paper, vector inequalities are understood componentwise.

[5] See Appendix Assumption A1 for detailed assumptions.

[6] To interpret types, choices, and outcomes as random variables and to define our mobility statistics as probabilities, we impose mild measurability conditions on primitives and induced objects (See Appendix Assumption A2).

does not create confusion, fix a reference generation t and suppress its time subscript, for example, write $z := z_t$ and $z_{-1} := z_{t-1}$.

Assume that the individual type θ has two components, $\theta = (h, w)$. The scalar $w \in \mathbb{R}_+$ represents innate ability. Assume ability is welfare-enhancing: more precisely, we assume that production functions and total resources F_z, F_k, T are non-decreasing, and one of them is increasing in w, so that the indirect utility v is increasing in talent.

The other component of heterogeneity is the vector $h_t \in \mathcal{H}$ recording dynastic history.[7] It encompasses the outcomes and welfare of prior generation, together with a time-invariant dynastic component $\chi \in \mathbb{R}_+$ (representing race, last name, etc.):

$$h_t = \big(x_{t-1}, y_{t-1}, z_{t-1}, k_{t-1}, v_{t-1}, \chi\big). \tag{2}$$

When moving from an individual model to a population study, we denote random variables with capital letters. For example, W_t denotes the random variable of talent and H_t family background in generation t. The induced professional outcome and welfare are the random variables $Z_t = z(H_t, W_t)$ and $V_t = v(H_t, W_t)$.[8]

The talent process $\{W_t\}_{t\geq 0}$ is exogenous and strictly stationary, with support on $\mathbb{R}_+$, and is independent of family background H. Its stationary marginal distribution F is strictly increasing and admits a smooth Lebesgue density f on $(0, \infty)$, with f bounded. We do not impose a time-homogeneous Markov structure. Instead, we discipline only the lag-1 dependence in talent through the copula C of (U_{t-1}, U_t), where $U_t = F(W_t)$. By Sklar's theorem, this copula is uniquely defined given continuity of F. We assume C is twice continuously differentiable on the interior of $(0, 1)^2$, which ensures well-behaved conditional distributions for W_t given W_{t-1}. We further assume that talent is *pairwise reversible at lag 1* if $(W_{-1}, W) \overset{d}{=} (W, W_{-1})$. Equivalently (under strict stationarity) the copula is symmetric: $C(u, v) = C(v, u)$ for all $u, v \in [0, 1]$.[9]

This distinction between innate ability w_t and family background h_t captures the idea that exceptional talent may arise in any family background, while the subsequent life course of such individuals is shaped by background h_t.[10]

[7] We assume $\mathcal{H}$ to be an open subset of $\mathbb{R}^6$ and endowed with the Euclidean topology. All technical assumptions are discussed and explained in the Appendix A.

[8] One technical assumption required later is the common image: For each $y \in \{z, v\}$, $y(h, 0) = 0$ for all $h \in \mathbb{R}$, and $\lim_{w\to\infty} y(h, w) = +\infty$ for all $h \in \mathbb{R}$. Equivalently, $\mathrm{Im}(y(h, \cdot)) = [0, \infty)$ for all h. See Appendix section A.2.3.

[9] This assumpti is a lag-1 symmetry and does not imply full time-reversibility of the entire process. It is only used in the proof of Theorem 1.

[10] Siblings share the same dynastic history, a feature we exploit in our empirical design. Moreover, the dynastic network contains inherent redundancy: for example, z and k jointly determine v. We maintain this general formulation deliberately, as it accommodates explicitly richer configurations such as $h = (z_{-1}, v_{-1})$.

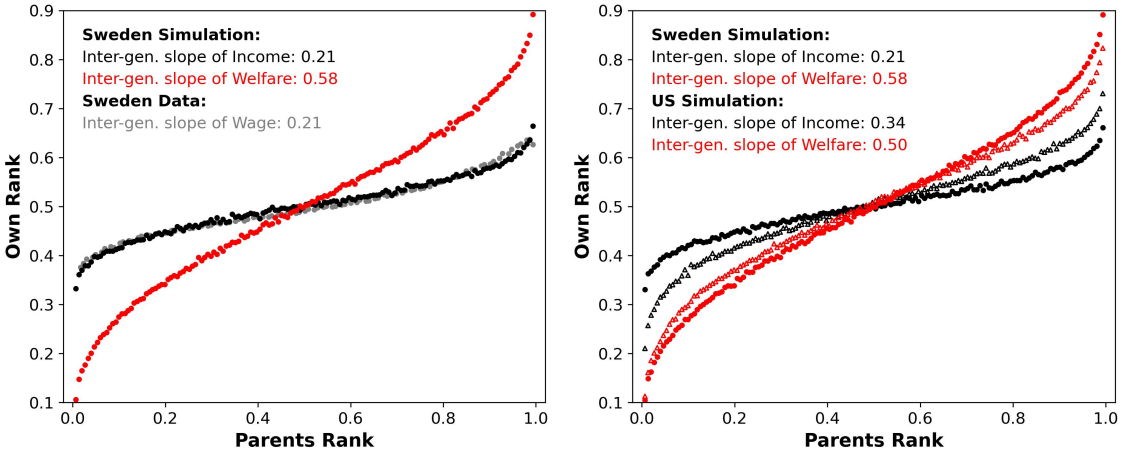
We illustrate the power of the setting with an example that incorporates the standard labor supply model but is augmented with a dynastic history.

Example: Labor supply model with history

Suppose the only relevant family background is parental career (similar to Atkinson (1980)), $h = z_{-1}$. The resource is one-dimensional, $L = 1$, and total resources do not depend on background, $T(\theta) = T$. Production functions are $F_z(x, \theta) = wxz_{-1}^{\vartheta}$ and $F_k(y, \theta) = yz_{-1}^{\beta}$, with $\vartheta, \beta > 0$. The limit case of $\vartheta = \beta = 0$ corresponds to the standard labor-supply model (absence of background privilege). The talent process is Markov, and (W, W_{-1}) is jointly log-normally distributed with correlation ρ.

Assume Cobb-Douglas utility over the two dimensions of life. In this case, we can analytically demonstrate that income mobility exceeds welfare mobility (see Appendix Section A.5.2 for details). For the more general Stone–Geary utility, we rely on simulations, and the results are presented in Figure 2. To highlight the influence of background factors, we impose a highly unequal career cost—a large value of β—and calibrate the intergenerational correlation of talent so that simulated income mobility aligns with the data (Panel a). We find that welfare mobility, as measured by the slope, is substantially lower than income mobility. When the career cost is less unequal, welfare remains more persistent than income, though the gap narrows.

Figure 2: Intergenerational mobility overestimates welfare mobility: A Simulation



(a) Simulation vs. Data: Country 1 (b) Country 1 vs 2 differing in ...

Note: This figure plots the intergenerational rank-rank conditional mean of income (black) and welfare (red). The DGP of talent is jointly log-normal with correlation ρ. Panel (a) compares a simulation of the model vs data. We set a high $\beta = 0.8$ for Sweden and chose $\rho = .21$ so that the implied income slope matches that from the data (depicted in gray). In Panel (b), the DGP of talent is the same across the two countries as before. Set $\beta = 0.095$ for the U.S., so the implied income slope matches that from data in Chetty et al. (2014). The number of observations in this simulation is 1.5m, similar to our data. see Section 1.1 and Appendix Section A.5.2 for details.

Applying the same talent process to the United States, we adjust the degree of unequal career so that the income mobility slope matches the American data (Panel b). Given that income persistence is higher in the US, this calibration requires a lower β than in Sweden. By construction, the US simulation assumes a more unequal cost of career. As a result, although income is more persistent across generations in the US, welfare persistence is lower than in Sweden. Strikingly, cross-country rankings depend on the metric: while upward income mobility is more common in Sweden, a comparable income gain in the US translates into a greater welfare improvement.[11]

1.1.1 Our vs. prior frameworks

Our framework nests many common models. Namely, it includes the standard labor-supply models, the models with altruism toward offspring, and models in which parents can invest in their children. As we saw in example above 1.1, our framework reduces to the standard labor–leisure formulation if heterogeneity is one dimensional $\theta = w \in \mathbb{R}_+$ (h is degenerate), input is one-dimensional time ($L = 1$), x denotes hours worked, y denotes leisure, and $F_z(x, w) = wx$ while $F_k(y, w) = y$. In this case, the professional life of a generation is income (consumption), and personal life is leisure. in another set of model with intergenerational linkage, income and consumption differ as there are inter-vivos transfers and investment in the next generation, denoted by I: preferences are over $\big(\bar{F}_z(\bar{x}, \theta) - I,\ \bar{F}_k(\bar{y}, I, \theta)\big)$. To embed this model, augment the resource vectors by defining $x := (\bar{x}, -I)$ and $y := (\bar{y}, I)$, and define encompassing production functions by $F_z(x, \theta) := \bar{F}_z(\bar{x}, \theta) - I$ and $F_k(y, \theta) := \bar{F}_k(\bar{y}, I, \theta)$. If $\bar{F}_k(\bar{y}, I, \theta)$ represents the welfare (or income) of the next generation, the model captures altruism.

A canonical framework in IGM literature (e.g., Becker and Chiswick (1966); Solon (1999)) assumes that generation-t preferences are over own consumption and the next generation's income, $\big(\mathrm{c}, \mathrm{y}_{+1}\big)$ where budget constraint is $\mathrm{c} + I = \mathrm{y}$ and income is result of own talent and parental investment: $\mathrm{y} = \mathrm{w} + \mathrm{R}\, I_{-1}$ where R is return to investment.[12] There is no intra-generation allocation decision in these models, only an intergenerational one. So our framework nests such models. To see how, let $L = 1$, set $F_z(x, \theta) := h + x$ and $F_k(y, \theta) := w + y$, $h = k_{-1}$, $w = \mathrm{w}_+$, $y = I$, $T(\theta) = T$. Although these models are encompassed by our framework (possibly with minor notational extensions as above), they modify the mapping between our theoretical objects and the empirical exercises; we return to this issue below.

[11] It is worth emphasizing that our assumption of a highly unequal cost of career in the US was chosen for illustrative, thought-provoking purposes; alternative calibrations are equally feasible.

[12] To avoid confusion with the main notation, variables in this model are written in upright Roman type.

1.2 Defining intergenerational mobility (IGM)

We define intergenerational mobility (IGM) as the probability that two dynasties reverse their relative positions across generations. In statistics, this is the probability of discordance (Schweizer and Wolff, 1981; Kochar, 2022; Nelsen, 2006). For example, professional-outcome IGM is

$$\mathcal{M}^z := \mathbb{P}\left(Z_t > \widetilde{Z}_t \,\middle|\, Z_{t-1} < \widetilde{Z}_{t-1}\right), \tag{3}$$

where $\{\widetilde{Z}_t\}$ and $\{Z_t\}$ are independent copies.[13] Throughout the paper, $\widetilde{X}$ denotes an independent copy of X. Appendix Section A.1 provides a detailed definition and useful derivation of this concept of IGM.

This definition has several virtues. (i) it depends solely on the copula of (Z_{t-1}, Z_t)—hence depends on correlation structure and is independent of inequality in each generation (marginal distributions). A copula-based mobility allows an ordinal comparison, which is essential given that one of the objects of interest here is welfare. (ii) it is inversely related to Kendall's rank correlation τ of (Z_{t-1}, Z_t), sc. $\mathcal{M}^z = \frac{1-\tau^z}{2}$. (iii) Most importantly, it is congenial for use in micro-founded theory, an advantage relative to Spearman rank correlation, popular in recent IGM studies.[14] We will come back to this point later.

1.3 Talent Function and MRS

The multidimensionality of dynastic history in (2) raises .... We first impose an order on dynastic histories. We interpret $h' \succ h$ ($h' \succeq h$) as dynasty h' being (weakly) more advantaged than h. We say h and h' are indifferent and write $h \sim h'$ when $h \succeq h'$ and $h' \succeq h$.

Assumption 1. *There exists a reflexive, complete, and transitive relation $\succeq$ on the set $\mathcal{H}$ of dynastic histories.*

The multidimensionality of dynastic history in (2) raises the question of how to operationalize the order in the analysis. We consider two approaches. The first works directly with the order $\succeq$ under weak regularity conditions, yielding simpler proofs under greater generality. The second imposes a dimensional-reduction assumption that maps h into a

[13] Under atomlessness, $\mathbb{P}(Z_{-1} < \widetilde{Z}_{-1}) = 1/2$, so $\mathcal{M}^z$ equals the unconditional probability of discordance.

[14] Kendall's τ has been used in only a few empirical studies of IGM (notable examples include Peters (1992); Althoff et al. (2025)), and its theoretical potential has received even less attention in the IGM literature. Outside the IGM context, it has been used to study assortative matching both theoretically (e.g., D'Agostino and Dardanoni (2009)) and empirically (e.g., Eika et al. (2019)).

scalar index, enabling differentiation and standard calculus; the key object is then the marginal rate of substitution (MRS) between talent and history. This approach is more intuitive and facilitates connections to empirical work.

1.3.1 Ordinal single-crossing assumption

When an individual from a less advantaged background overtakes a more advantaged peer in the personal-life dimension, she also does so in the career dimension. This concept is formalized in the following assumption.

Assumption 2 (Milgrom–Shannon single crossing)**.** *We say that k single-crosses z if, for $h' \succ h$,*

$$k\left(h,w\right) \geq k\left(h',w'\right) \;\Rightarrow\; z\left(h,w\right) > z\left(h',w'\right). \tag{4}$$

Assumption 2 is a *strict* version of Milgrom–Shannon single-crossing condition (Milgrom and Shannon, 1994). To see the connection, define the lexicographic order among θ—namely, $(h,w) \succeq (h',w')$ if either $h \succ h'$ or hh' and $w \geq w'$— and the real function $\varphi(h,w,t)$, so that $\varphi(h,w,1) = z(h,w)$ and $\varphi(h,w,0) = k(h,w)$. The Assumption 2 is a strict single-crossing property of $\varphi(h,w,t)$. This connection helps us soon to appeal to Theorem 3 of (Milgrom and Shannon, 1994). Moreover, note that Assumption 2 implies single crossing of v over z; see Appendix Section A.3.6 (Proof Z2).

We can write a more useful version of the Assumption 2 using the talent metric function: the minimum level of talent required to overcome a given disadvantage in background. We formalize it as follows.

Definition 1. *Define the talent metric function μ^y for $y \in \{z,k,v\}$ as the talent level required to achieve the same outcome under history h' as under (h,w):*

$$y\left(h,w\right) = y\left(h',\mu^y\left(h';h,w\right)\right), \tag{5}$$

As $y \in \{z,k,v\}$ are strictly increasing in w, their talent metric function exists and is unique. It can also be represented as $\sigma^v(h', v\left(h,w\right))$, where σ^v is the the implicit inverse function: $v\big(h,\sigma^v(h,v)\big) = v$. This representation — i.e., $\sigma^v(h', v\left(h,w\right))$— parallels the money-metric expenditure function $e(p', v\left(p,w\right))$.[15] Intuitively, the former measures the additional talent required to compensate for a difference in background, and the latter measures the additional income required to compensate for a price change.

[15]This representation also shows that talent metric is a separable function,i.e., $\partial_h \mu^y / \partial_w \mu^y$ is independent of h'.

Claim 1. *Milgrom-Shannon single crossing (Assumption 2) is equivalent to the following ranking of talent metrics: for any $w \in \mathbb{R}_+$ and any histories h, h',*

$$h' \succ h \;\Rightarrow\; \mu^z(h'; h, w) > \mu^k(h'; h, w)\,, \qquad h' \prec h \;\Rightarrow\; \mu^k(h'; h, w) > \mu^z(h'; h, w)\,.$$

Proof in Appendix Section A.3.1. One interpretation of this claim is that the talent metric function as a function of h crosses zero only once at $h = h'$. [16]

Equivalently, Assumption 4 and Claim 1 can be stated in terms of career and welfare.

1.3.2 Differentiability and Spence–Mirrlees single crossing

We can replace the Milgrom-Shannon single crossing (Assumption 2) with a much stronger, but more intuitive, assumption when we are willing to suppose that the multi-dimensionality of h can be reduced.

Assumption 3 (Total ranking and outcome invariance)**.** *(i) The relation $\succeq$ is continuous, locally nonsatiated, and the topological boundary of each upper contour set is a C^2-manifold in $\mathcal{H}$ for every $h \in \mathcal{H}$. (ii) For each outcome $y \in \{z, k\}$ and two indifferent background $h, h' \in \mathcal{H}$, $h \sim h'$ implies that $y(h, w) = y(h', w)$ for all $w \in \mathbb{R}_+$.*

Appendix Claim A1 shows that the first part of the assumption allows us to appeal to Proposition 2.3.9 in Mas-Colell (1985) that guarantees the existence of a C^2 utility index representing $\succeq$: $h' \succeq h \Leftrightarrow \beta(h') \geq \beta(h)$ for a real function $\beta \in C^2$. The second part, then, helps us to define functions $\hat{y}$ for $y \in \{c, k\}$ so that $\hat{y}(b, w) := y(h, w)$ when $\beta(h) = b$. $\widehat{y}$ is second continuously differentiable and increasing in both elements. $\widehat{y}_b > 0$ as a result of total ranking assumption 3 and $\widehat{y}_w > 0$ since $g_w > 0$. [17] Analogous statements apply to v, given its dependence on c, k.

Assumption 4 (Spence–Mirrlees single crossing)**.** *The marginal rate of substitution is higher for career than family, that is, for any (b, w),*

$$\Delta := MRS^z - MRS^k := \frac{\partial_b \hat{k}(b, w)}{\partial_w \hat{k}(b, w)} - \frac{\partial_b \hat{z}(b, w)}{\partial_w \hat{z}(b, w)} > 0 \tag{6}$$

The name of this assumption coming from its equivalence to Spence–Mirrlees' single crossing of the function φ (Laffont and Martimort, 2002; Bolton and Dewatripont, 2005).[18]

[16] In the simple case of one-dimensional history, these two conditions can be written concisely as $[h' - h] \times \left[\mu^z(h'; h, w) - \mu^k(h'; h, w)\right] > 0$.

[17] For more details of these steps see Appendix Claim A1.

[18] The Spence-Mirrlees (single–crossing) condition is often stated as a positive cross-derivative under quasi-linearity (Salanié, 2005).

$MRS^{\hat{z}} > MRS^{\hat{v}} \Leftrightarrow MRS^{\hat{z}} > MRS^{\hat{k}}$. See Z2 for proof in Appendix Section A.3.6.

A simple way to see the connection between the MRS and talent metric is to differentiate the equivalent to (5) under Assumption 3 — that is, $\hat{z}\left(b,w\right) = \hat{z}\left(\widetilde{b}, \mu^{\hat{z}}(\widetilde{b};b,w)\right)$— with respect to $\widetilde{b}$, yields

$$\frac{\partial}{\partial \widetilde{b}} \mu^{\hat{z}}\left(\widetilde{b};b,w\right) = MRS^{z}$$

where the MRS is at $(\widetilde{b}, \mu^{\hat{z}}(\widetilde{b};b,w))$.

Claim 2. *Under assumptions 3, two single crossing conditions (Assumptions 2 and 4) are equivalent.*

We provide two proofs of this claim. One direct, and one appeals to Theorem 3 of (Milgrom and Shannon, 1994). See Appendix Section A.3.3. Given this equivalence, we provide our analysis using Milgrom-Shannon sign crossing (Assumption 2), but we can interpret them as MRS under Assumption 4.

1.4 Welfare vs. Income IGM

Given this order 1, we define *Fundamental-mobility* as the probability that a dynasty from a less advantaged history overtakes one from a more advantaged history (the discordance probability of outcome and background). For example, the fundamental-mobility of career is:

$$\widehat{\mathcal{M}}^{z} := \mathbb{P}\Big(Z_t > \widetilde{Z}_t \,\Big|\, \widetilde{H}_t \succ H_t\Big), \tag{7}$$

where (H_t, Z_t) and $(\widetilde{H}_t, \widetilde{Z}_t)$ are independent copies. Similarly define $\widehat{\mathcal{M}}^{k}$ or $\widehat{\mathcal{M}}^{v}$.

Fundamental-mobility is more fundamental than standrad mobility $\mathcal{M}^{z}$, but it is typically not directly observable. It is not often directly observable because H is high-dimensional and only partially observed in the data. In the special case in which the relevant component of dynastic history is parental career, $h = z_{-1}$ (similar to the example of Section 1.1), Fundamental-mobility coincides with mobility: $\widehat{\mathcal{M}}^{z} = \mathcal{M}^{z}$.

The next result shows that if welfare exhibits a higher MRS than career, $\mathrm{MRS}^{v} > \mathrm{MRS}^{z}$, then any welfare climber is also a career climber: in a random pair, if the initially disadvantaged individual overtakes in welfare, she also overtakes in career.

Proposition 1. *Single crossing condition of v over z implies Fundamental-mobility of career exceeds that of welfare, that is $\widehat{\mathcal{M}}^{z} > \widehat{\mathcal{M}}^{v}$. Vice versa, when $\widehat{\mathcal{M}}^{z} > \widehat{\mathcal{M}}^{v}$ holds for any history and talent DGP, the Single crossing condition of v over z.*

The proof under the milder assumption is straightforward, ... The proof is longer under path 1. See A.3.4 for both proofs.

Corollary 1. *Assume $h = z_{-1}$. Then the single-crossing condition of v over c implies that the mobility of career exceeds that of welfare.*

Corollary 2. *Assume the absence of talent and that dynastic history has a higher concordance with welfare than with income, then the Single crossing condition for welfare over career implies that career mobility exceeds welfare mobility.*

Proof in Appendix Section A.3.5.

Assumption 5. *For any history pairs $h, \widetilde{h}$ with $h \neq \widetilde{h}$, the signle-crossing gap defined by Assumption 2 strictly widens with talent; that is, $|\mu^v - \mu^z|$ is strictly increasing in w on $\mathbb{R}_+$.*

Under Total Ranking, the Widening Gap assumption Assumption 5 is equivalent to $\partial_w \Delta > 0$.

One technical assumption that we use in the proof of next proposition states that, conditional on being exactly tied in the parent generation, the next generation's comparison is systematically tilted in favor of the advantaged background by a margin bounded away from a fair coin flip. This assumption, in the case of welfare, loosely speaking, is

$$\mathbb{P}(\widetilde{V} > V \mid \widetilde{V}_{-1} = V_{-1}, H = h < \widetilde{H} = \widetilde{h}) > \tfrac{1}{2}. \tag{8}$$

We will assume this for all outcomes, more specifically, for outcomes between income and welfare. We will relegate the details of this assumption to the Appendix (see Assumption A4).

Theorem 1. *If family background is time-invariant, and under sinle-crossing and widening gap Assumptions (2 and 5), then mobility of career exceeds that of welfare, that is, $\mathcal{M}^z > \mathcal{M}^v$*

The idea of the proof is as follows: Mobility can be expressed as a functional of the talent metric, $\mathcal{M}^y = 2\mathbb{E}_{H,\widetilde{H}}\left[\bar{\mathcal{J}}(\mu^y_{H,\widetilde{H}})\right]$. Assumption 5 implies that the income talent metric can be obtained from adding Dirac functions to the welfare talent metric. The issue is that this "needle" path is not differentiable. Step 1 implements this using a differentiable path using a smooth cutoff function. Lemmas A1 and A2 then provide (pathwise) a Gateaux derivative representation and the regularity needed to differentiate this functional along the path. Steps 2 and 3 show that adding the functional decreases along the path, under talent time invariance and Assumption A4. Step 4 establishes the theorem by tracing the functional (mobility) along the path from income mobility to welfare. Full Proof is Appendix Section A.4.

We now present a special case of the previous theorem in a parametric setting, where the proof is considerably simpler. We then use this framework to study, through simulations, several cases with time-varying background conditions.

Claim 3. *Assume a time-invariant Bernoulli background, a Markovian joint log-normal talent,i.e.,* $(\ln W_{-1}, \ln W_t) \sim \mathcal{N}(0,0;1,1;\rho)$ *and that talent-metric functions are linear. Then* $\mathcal{M}^z > \mathcal{M}^v$.

Proof relying on characteristics of normal distribution in Appendix Section A.5.3.
Few simulations

$$Z_t = W_t 1_{H=0} + a^z W_t 1_{H=1}$$

We use simulations for two additional cases where
first case: H_t is Bernoulli that is 1 if Z_{t-1} is above median.
second case: H_t is Bernoulli^2 where the first one is Z_{t-1} is above median and the second one is V_{t-1} is above median.

2 From Theory to Data: An Empirical Test

A key empirical challenge is that neither innate talent w_t nor (period) utility v_t is observed. In the data we instead observe (i) proxies for background characteristics b (e.g., parental income) and sibling design, (ii) some career outcomes z (e.g., hours worked, earnings trajectories), and (iii) a few family outcomes k (e.g., fertility). We also observe selected components of the input vectors x_t and y_t (e.g., hours, parental leave).

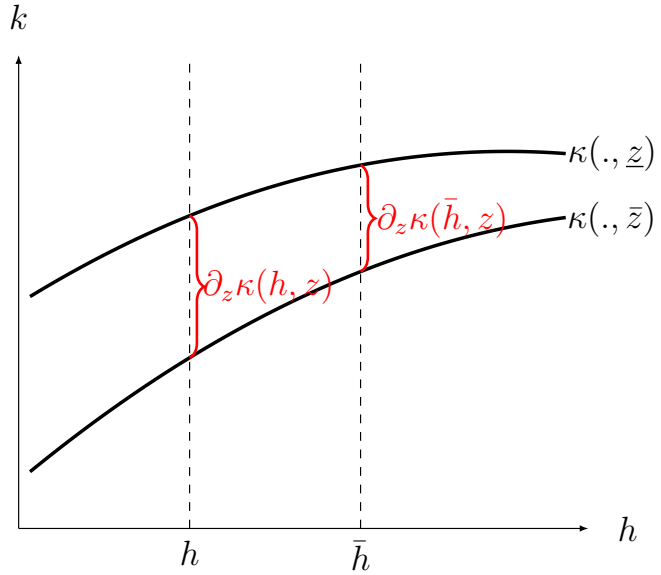
2.1 Empirical Test: Idea

Our empirical object is the background-specific *family cost of career*: the within-family difference in k associated with a difference in career outcomes z, holding background fixed. More formally, the observable iso-career lines are defined as

$$\kappa(h,z) = k\big(h, \sigma^z(h,z)\big)$$

where like before σ^z is the the implicit inverse function: $z\big(h, \sigma^z(h,z)\big) = z$. Let us illustrate how our empirical test use iso-career lines with sister comparisons in Figure 3.

Figure 3: Test the Main Assumption 5 using Observational Data



Note: This figure plots iso-career lines in family-background space, that is, family outcome k against background b for two sibling pairs: one pair has background h and another $\bar{h}$. Both families have one sibling with career $\bar{z}$ and another with career $\underline{z}$.

Consider two sisters who share background h but differ in talent. The talent difference induces different career outcomes $\underline{z}$ and $\bar{z}$. The corresponding difference in family outcomes, the family cost of career for this family, is

$$\partial_z\kappa(h,z) = \partial_w k(h,w)\,\sigma_z^z(h,z) = \frac{\partial_w k(h,w)}{\partial_w z(h,w)}, \tag{9}$$

evaluated at $w = \sigma^z(h,z)$. The same family cost of career for sibling-pair in another family with slightly different background is $\partial_{zh}\kappa(h,z)$ which can be written as

$$\partial_{zh}\kappa(h,z) = \frac{\partial_w k}{\partial_w z}\,\Delta\left(\frac{\partial_{ww}k}{\partial_w k} + \frac{\partial_w \Delta}{\Delta}\right), \tag{10}$$

with right hand side evaluated at $(h,\sigma^z(h,z))$ and Δ is the differnce in MRS as defined in **??**.[19] Therefore, under Assumption 1 ($\Delta > 0$):

$$\partial_{hz}\kappa(h,z) > 0 \implies \frac{\partial_{ww}k}{\partial_w k} + \frac{\partial_w\Delta}{\Delta} > 0 \implies \partial_w\Delta > 0$$

This is the local (infinitesimal) version of Assumption 5.

Differentiating with respect to h gives: $\partial_h\kappa(h,z) = \partial_w k\ \Delta$, where again with right hand side is evaluated at $(h,\sigma^z(h,z))$. This implies that the Assumption X is eequivalent of increasing iso-career lines.

[19] For the derivation see A.3.6.

2.2 Empirical Test: Implementation

We empirically test the unequal cost of a career, $\partial_{hz}\kappa(h,z) > 0$, by regressing the lifetime family outcome of an individual, k, on the individual's lifetime income, z, parental lifetime income, z_{-1}, and the interaction of these two elements. In this way, parental lifetime income stands as a proxy for background. We will later augment this with a within-sibling comparison using family fixed effects.

More specifically, we run the following regression:

$$k = \mathbf{f}(z) + \mathbf{g}(z_{-1}) + \sum_{i=1,\dots,9} \beta_i \underbrace{\mathbb{1}(\mathbf{D}(z) = 10)}_{\text{Own Income Top Decile}} \times \underbrace{\mathbb{1}(\mathbf{D}(z_{-1}) = i)}_{\text{Parental Income Deciles}} + \varepsilon \tag{11}$$

where $\mathbf{D}(.)$ is a decile function and $\mathbf{f}(.)$ and $\mathbf{g}(.)$ are flexible functions. We implement this by flexibly controlling for individual and parental lifetime income using gender-specific dummies of rank percentiles. $\mathbb{1}(\mathbf{D}(z) = 10)$ then indicates having a lifetime income in the top decile which we refer to as "having a career".

Our coefficients of interest are $\{\beta_i\}$, which capture the interaction between having a top-decile income ("having a career") and parental income deciles. When we interact the career indicator with indicators for parental lifetime income deciles we omit the highest decile as the reference group. The coefficients β_i therefore measure, for each parental income group i, the difference in family outcomes between individuals with and without a career, relative to the corresponding difference for individuals from top-decile parental backgrounds. In this way, our design captures a difference-in-differences parameter, albeit without any causal interpretation. A negative coefficient implies that individuals from lower-income backgrounds experience larger reductions in family outcomes when attaining careers, relative to individuals from high-income backgrounds. Thus, β_i captures how the family-related cost of upward mobility varies with parental background. We place particular emphasis on β_1, which compares individuals from bottom-decile parental backgrounds to those from the top decile.

We complement this specification by running the regression using a more parametric approach. To do so, we replace the indicator of having a career with a linear rank measure in the interaction term:

$$k = \mathbf{f}(z) + \mathbf{g}(z_{-1}) + \sum_{i=1}^{9} \beta_i \underbrace{\mathbf{R}(z)}_{\text{Own Income Rank}} \times \underbrace{\mathbb{1}(\mathbf{D}(z_{-1}) = i)}_{\text{Parental Income Deciles}} + \varepsilon \tag{12}$$

where $\mathbf{R}(.)$ describes a rank function and $\mathbf{R}(z)$ captures moving up in rank.

Finally, we add a series of controls to Regression 11. We flexibly control for family

fixed effects, spouses' lifetime income percentiles, age at first birth fixed effects, and years of education fixed effects. Including father fixed effects mirrors our initial exercise of comparing two sisters in Figure 1.

$$k = \mathbf{f}(z) + \delta(h) + \sum_{i=1,\ldots,9} \beta_i \underbrace{\mathbb{1}(\mathbf{D}(z) = 10)}_{\text{Own Income Top Decile}} \times \underbrace{\mathbb{1}(\mathbf{D}(z_{-1}) = i)}_{\text{Parental Income Deciles}} + \varepsilon \tag{13}$$

2.3 Discussion

Connecting back to the career-family trade-off and unequal career cost. Similar to converting firms' profit maximization into cost and scale optimization in production theory, we reformulate the individual problem into a two-step optimization. In the inner problem, the individual maximizes her personal life given the amount of professional life that she has to achieve, that is,

$$\mathbf{k}(z, \theta) := \max_{x_z, x_k} F_k(x_k, \theta) \tag{14}$$

under the career constraint $F_z(x, \theta) \geq z$ and the resource constraint $x_z + x_k \leq r(\theta)$. So, the function $\mathbf{k}$ is the optimal personal life given a minimum level of professional life. It is decreasing, $\frac{\partial \mathbf{k}}{\partial z} < 0$, where the derivative captures the loss of personal life for one unit of career gain. The outer problem is balancing the trade-off between professional and personal life, that is our main focus:

$$v(\theta) = \max_{z} u(z, \mathbf{k}(z, \theta)) \tag{15}$$

The gain from the two-step formulation is that background influence is entirely captured in the personal-life function $\mathbf{k}$. The function summarizes the two production functions and resource constraints.

This concise two-step representation of our micro-founded model is helpful to investigate two important issues. First, within this framework, unequal costs of care can be defined from an individual perspective as $\partial_{zh}\mathbf{k}(z, h, w) > 0$: The family cost of career $\partial_z \mathbf{k}$ from the individual perspective is lower for high-bron people. Importantly, this object is defined off equilibrium, even though it is highly relevant from the standpoint of individual decision-making. By contrast, the notion of unequal costs of care that we examine in the empirical analysis— and that provides a test of the theory's main assumption—is an equilibrium object. Specifically, it also corresponds to a positive cross derivative with respect to z and h, but of a different function that does not condition on w (or talent), namely $\partial_{zh}\kappa(z, h) > 0$. In other words, the former holds talent fixed, whereas the latter does not. This distinction is advantageous in observational data, where innate talent is

not observed. In retrospect, it is not surprising that mobility is related to a population property rather than an individual-level feature that is usually relevant as they govern comparative statistics and behavioral responses.

The second issue is an important dimension that we have abstracted from so far: preference heterogeneity. Returning to Figure 1, a natural question is whether the pattern documented there could be generated solely by heterogeneity in preferences. The answer is no: preference heterogeneity alone cannot reproduce the pattern. Intuitively, allowing preferences to vary across individuals can change the marginal distributions of career and family outcomes, but it does not determine how these outcomes co-move. Put differently, preference heterogeneity affects the marginal distributions, whereas the empirical pattern in Figure 1 is about the joint dependence structure (the copula) linking career and family outcomes. To see this, return to equation (15) and allow utility to vary across individuals:

$$\max_{z} \; u(z, \mathbf{k}(z, \theta), \tau) . \tag{16}$$

If θ is absent, heterogeneity in the taste shifter τ—regardless of its data-generating process, whether it has hereditary components, whether its variance varies with family background, and so on—can only operate by changing the distribution of the optimal career choice z, and thus, through $\mathbf{k}(\cdot)$, the marginal distribution of family outcomes. What it cannot do is change the dependence between career and family outcomes: variation in τ moves individuals along the same $\mathbf{k}(\cdot)$ mapping, while the joint dependence is governed by the functional form of $\mathbf{k}(\cdot)$ (and, when present, by how θ shifts that mapping). Therefore, the pattern we observe implies that $\mathbf{k}$ must depend on family background.

2.4 Data and Estimation Sample

We use individual-level data on the universe of the Swedish population from several administrative registers on (i) labor, capital, and business income spanning from 1968 to 2020 from the Income tax database, (ii) family relations from the Multi-generational register, (iii-iv) demographic information from the longitudinal integrated database for health insurance and labour market studies (*LISA*) and the register of the total population (*RTB*) spanning from 1990 to 2020 for the former and 1968 to 2020 for the latter, (v) cognitive and non-cognitive skills from Swedish draft data, (vi) firm characteristics from Employer-employee matched dataset, and (vii) test scores at high school from the Educational register. We complement these datasets with (viii) individual-level survey data on wages and hours from the Wage Structure Survey and (ix) survey data on working conditions from the Work Environment Survey (AMU).

We construct several variables. Individual lifetime income, z, is measured as the sum of labor income from ages 25 to 45. We rank incomes within the individual's birth cohort. If individuals have the same income level, we assign the group's average rank, following Chetty et al. (2014). Parental income rank, z_{-1}, is measured as the sum of the total income of the mother and the father. Total income contains labor income but also business and capital income. We sum parental income over the individual's childhood, i.e., the ages 0 to 20.[20] Then, we rank incomes within the birth cohort of the individual following Chetty et al. (2014). We assign the average rank of the group in case of the same income levels.

We measure personal life outcomes through family-related indicators. Our primary focus is on the number of children a person has in their lifetime. This is quantified as the number of biological children born before age 45, which serves as a proxy for the Completed Fertility Rate. We refer to it as a proxy because the analysis is restricted to age 45 to incorporate a broader range of cohorts into the study. Unlike the commonly used Total Fertility Rate, which reflects the average number of children a woman would have if she experienced the current age-specific fertility rates throughout her reproductive years, the Completed Fertility Rate directly measures the total number of children born to an individual by the end of her reproductive span. We complement this measure of the number of children with additional family-related outcomes, such as age at first birth, age at marriage, and average days of parental leave per child.

We restrict the analysis to individuals born in Sweden between 1955 and 1975. This restriction enables us to construct a balanced sample of individuals who reached at least age 25 by 1968 (the first year of income data) and 45 by 2020 (the last year of the data). We exclude individuals who die before the age of 45, have non-identifiable biological parents, or have parents born outside of Sweden. Furthermore, we drop individuals who receive any disability insurance or reside outside Sweden between the ages of 35 and 45.[21] This data restriction results in 1,589,961 individuals. In Table 1, we report sample characteristics of our main variables. On average, individuals have 1.9 children before age 45, are 29 years old at the first birth, and take, on average, 213 days of parental leave per child.

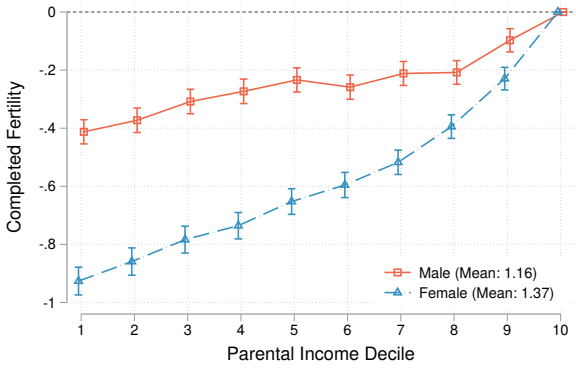
[20] If individuals are born outside the period for which we have data, we use a shorter period. For example, for the cohort born in 1967, we use ages 1 to 20; for the cohort born in 1966, we use the ages 2 to 20, ..., up to the cohort born in 1955, for which we use ages 13 to 20.

[21] We start at age 35 due to data restrictions of disability insurance and residence status, which are variables of the LISA dataset beginning in 1990.

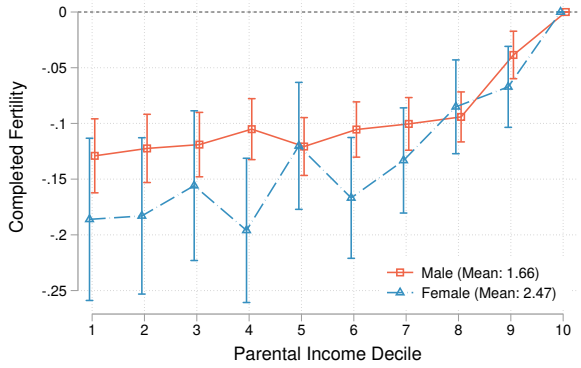
3 Unequal Cost of Career: Empirical evidence

This section presents empirical evidence on the unequal cost of career across parental backgrounds. We examine completed fertility, timing of family formation, and time spent with children, proxied by parental leave days per child.

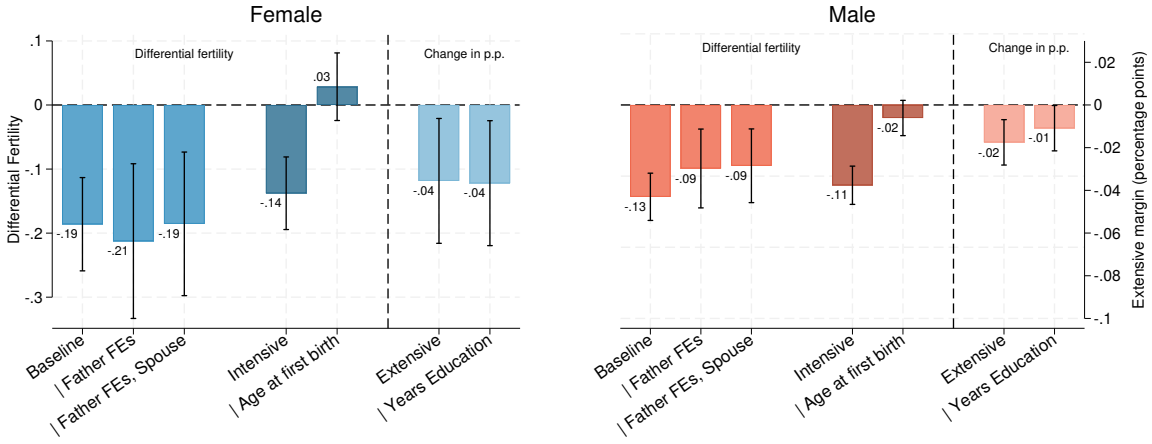
Figure 4: Family Cost of Upward Mobility: Completed Fertility



(a) Linear Specification, regression 12



(b) Non-parametric Specification, regression 11



(c) Extensive/Intensive Margins + Controls

Note: This figure illustrates the difference in fertility between individuals at the top vs. the bottom of lifetime income at each parental-income decile compared with the same difference for the top-decile parents (the reference group). It plots the coefficients of the interaction terms, $\{\beta_i\}$, in the regression of fertility on gender-specific non-parametric functions of their own lifetime income z and their parental lifetime income z_{-1}. The interaction terms has linear specification Panel 4a and a non-parametric specification in Panel 4b. The latter is:

$$k = \mathbf{f}(z) + \mathbf{g}(z_{-1}) + \sum_{i=1,\ldots,9} \beta_i \underbrace{\mathbb{1}(\mathbf{D}(z) = 10)}_{\text{Own Income Top Decile}} \times \underbrace{\mathbb{1}(\mathbf{D}(z_{-1}) = i)}_{\text{Parental Income Deciles}} + \varepsilon$$

In Panel 4c, we report only β_1 from the regression above where we add controls as indicated in the panel. The first 5 bars measure the difference in the number of children—the *differential fertility* (left axis)—while the next two bars capture differences in the extensive margin measured in percentage points (right axis). The baseline mean of Panel 4a and 4b is calculated as the average completed fertility among individuals in the bottom decile of income with bottom-decile-income parents. The estimation sample comprises 1,589,961 individuals born in Sweden between 1955 and 1975, with parents also born in Sweden. *Completed fertility* is the number of biological offspring born before age 45.

3.1 Completed Fertility Rate

Figure 4 displays the cost of career measured by the lifetime number of children (Completed Fertility Rate). Figures 4a and 4b use regression specification 12 and 11, respectively. Coefficients are normalized to zero for individuals with top-decile parental income, so that negative values indicate a greater fertility shortfall as income increases for individuals from disadvantaged backgrounds relative to their affluent counterparts.

Under the linear interaction specification (Panel 4a), extrapolation from the bottom to the top of the own-income distribution reveals a pronounced parental gradient for women: the estimated differential is approximately -0.93 children for women from parental decile 1 (relative to zero for parental decile 10). The corresponding estimate for men is markedly smaller (around -0.42), albeit similarly precise. Accordingly, linear extrapolation implies that the "unequal family cost" of attaining top-income status is more than twice as large for women as for men.

The nonparametric specification (Panel 4b), which compares top-decile own income with the bottom 90 percent, yields the same qualitative pattern but with attenuated magnitudes. For parental decile 1, the differential is approximately -0.19 for women and -0.13 for men, and it converges to 0 by parental decile 10. The gender gap, therefore, remains, but it is noticeably smaller than in the linear specification. Standard errors for women are typically larger in this panel, consistent with fewer women at the top of the own-income distribution.

Figure 4c decomposes the outcome to extensive and intensive margins, using the specification 11. The patterns are similar across the extensive and intensive margins, with the magnitude of the effects indicating a comparable contribution from each margin. The differential impact by gender is largely attributable to differences in the extensive margin.

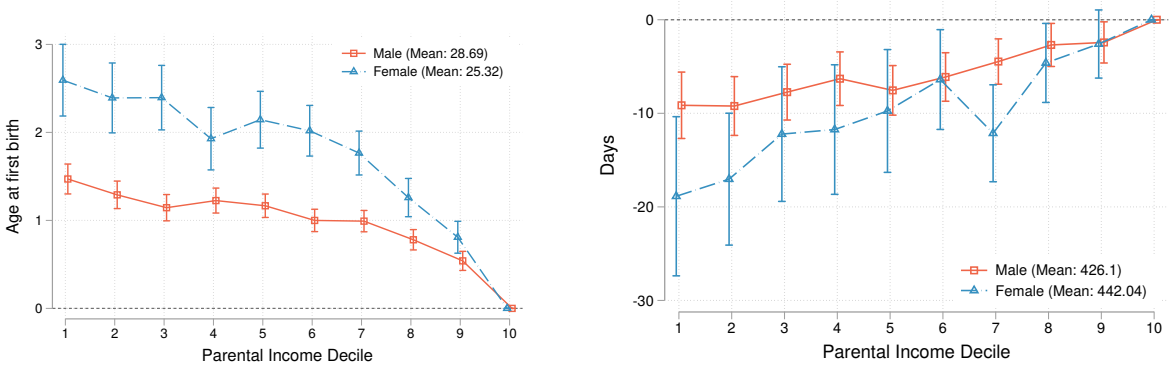
Figure 4c also explores potential mechanisms and alternative approaches to accounting for parental background. Substituting flexible parental-income controls with father fixed effects yields comparable magnitudes for women and a somewhat smaller, yet still negative, estimate for men, with fixed-effects estimates identified from within-family variation.[22] By adding father (or mother) fixed effects, we control not only for parental lifetime income and within-family preferences for personal life but also for additional unobserved factors that are constant across siblings, such as parents' business networks or parental educational investments in the children.

If individuals face a trade-off between time spent with their family and a career, the spouse's career matters, especially when raising children is done within the household.

[22] We run complementary regressions using the biological mother to identify siblings within one family.

Augmenting the model with spouse income percentiles leaves the estimates essentially unchanged, indicating that differences in partner resources do not principally account for the fertility gradient. We will return to the spouses and study assortative mating in Section 5.2. Lastly, conditioning on age at first birth largely absorbs the remaining intensive-margin differential, consistent with the timing of family formation serving as a key mediating factor—thus motivating the next section's focus on fertility timing.

Figure 5: It's About Time: Timing of Birth and Parental Leave



(a) Age at First birth

(b) Parental Leave per Child

Note: This figure illustrates the difference in the age at first birth (Panel 5a) and days on parental leave (Panel 5b) between individuals at the top vs. the bottom of lifetime income at each parental-income decile compared with the same difference for the top-decile parents (the reference group). It plots the coefficients of the interaction terms, $\{\beta_i\}$, in specification 11. *Parental leave per child* is the total number of days parents are on parental leave per child. In Appendix Figure C3, we decompose days spent on parental leave into extensive and intensive margins and report days separately for own and spouse's parental leave.

3.2 Delayed Family Formation and Parental Leave

Figure 5 demonstrates that family formation is substantially postponed among individuals from disadvantaged backgrounds who ascend to the top of the lifetime income distribution. Employing the same specification as regression 11, Panel A considers age at first birth as the outcome variable. The parental background gradient is especially pronounced for women: those from low-income families who achieve top-decile own income are, on average, more than 2 years older at the time of their first child's birth relative to their low-income peers, once we compare them with their affluent-origin peers. The corresponding effect for men is smaller but remains economically meaningful.

This differential reflects a shift in the timing of first births, rather than a uniform postponement across all ages. Compared to their affluent-origin counterparts, upwardly mobile individuals from disadvantaged backgrounds are substantially less likely to have a first child between ages 18 and 25, and markedly more likely to do so after age 35 (Appendix Figure C1 and C2). This pattern aligns with prolonged educational investment, a

mechanism explored in the subsequent section.

Panel B of Figure 5 examines parental leave taken per child. Here as well, the parental-background gradient is marked: upwardly mobile individuals from disadvantaged backgrounds, together with their spouses, take approximately 10 to 20 fewer days of parental leave per child than otherwise comparable individuals from affluent origins. The reduction is more pronounced for women than for men. Appendix Figure C3 indicates that this disparity is driven predominantly by female leave within the household.

This gap is particularly striking in the Swedish context, where parental leave is universally accessible via the social insurance system. Nevertheless, collective agreements and employer-provided top-ups introduce heterogeneity in effective compensation during leave, which may serve to reinforce socioeconomic disparities in leave-taking behavior.

4 Disadvantages and Mechanisms

The unequal cost of a career comes from two contrasting paths to success. In this section, we first document differences in barriers to success, namely time disadvantages and location barriers. We then analyze the mechanisms through which these disadvantages materialize. That is, we compare educational and professional pathways between selfmades and privileged achievers, i.e., achievers from high vs. low socio-economic backgrounds.

4.1 Disadvantages Across Time and Space

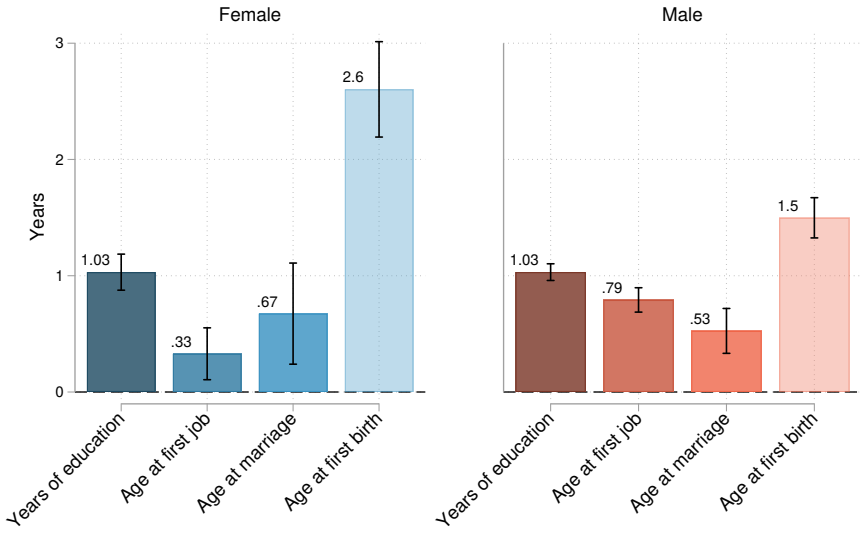
Selfmades face barriers to success across time. Figure 6 shows that to reach the top decile of the income distribution, individuals from bottom decile backgrounds study one year longer. This additional year of studies carries over into the timing of the professional and family life. Specifically, we find that disadvantaged achievers start their first full employment on average 0.4 to 0.8 years later and postpone both marriage and childbirth.

Besides drawbacks in time, disadvantaged achievers further experience drawbacks across space. Figure 7 demonstrates that underprivileged achievers geographic mobility separates them from their families. We measure this twofold: First, the left bar indicates they are 0.1 percentage points more likely to leave their birth county. Second, they live farther apart from their parents later in life. In particular, at age 35 individuals who have a career and come from bottom-decile parents live on average about 30 km (47 km for females) farther away from their parent's residence compared to individuals from high-income parents. Finally, we investigate from which municipalities they are mov-

ing. Appendix Figure C10a reveals that selfmade achievers are born into higher-ranked municipalities. One interpretation is that in order to achieve the same career as their counterparts from high-income parents selfmade achievers need a high-income environment. One specific aspect of this argument—networks—is discussed in the following section (see Figure 11).

This suggests that individuals from less advantaged backgrounds often need to exhibit greater time and geographic flexibility, sometimes at the cost of leaving their hometowns to secure career advancement. Nakamura et al. (2022) suggest a significant disutility of geographical moves. This is another aspect of personal life affecting welfare mobility.

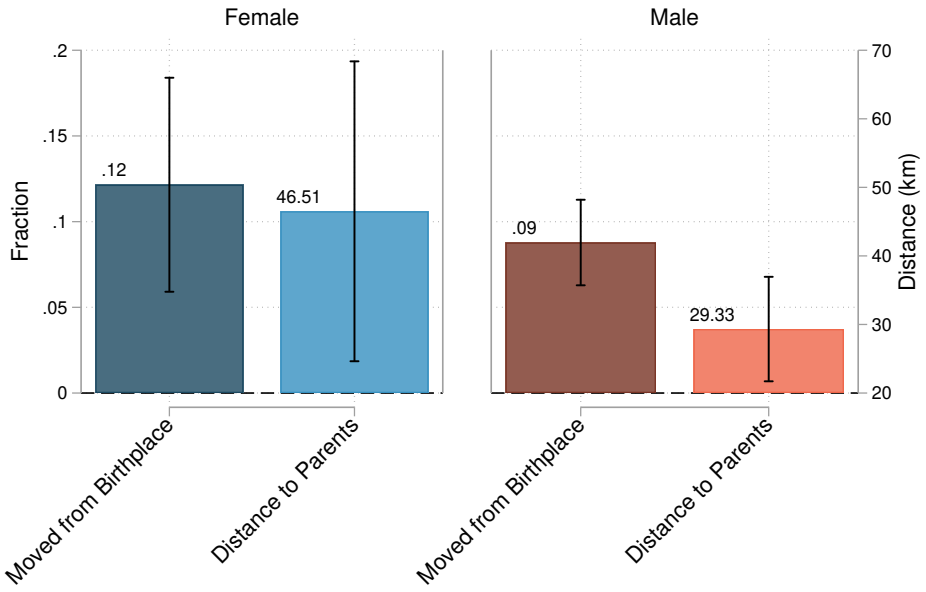
Figure 6: Time Drawbacks: Length of Studies, Start in the Labor Market, Start of a Family



(a) Age at starting Family and Professional Life

Note: This figure illustrates the difference in the timing of family and career outcomes between individuals at the top vs. the bottom of lifetime income at each parental-income decile compared with the same difference for the top-decile parents. It plots the coefficient of the interaction term, β_1, in the regression of the outcome on gender-specific nonparametric functions of their own lifetime income and their parents' lifetime income, specification 11. *Age at marriage* is conditional on being married before the age of 45. *Age at first child* is conditional on becoming a parent before age 45. More details are provided in Section 2.

Figure 7: Location Drawbacks: Geographical Mobility



Note: This figure illustrates the difference in the geographical mobility measures between individuals at the top vs. the bottom of lifetime income at each parental-income decile compared with the same difference for the top-decile parents. It plots the coefficient of the interaction term, β_1, in the regression of the outcome on gender-specific nonparametric functions of their own lifetime income and their parents' lifetime income, specification 11. *Moved from birthplace* describes an indicator that takes the value one if the individual's county of residence at age 45 differs from the county of birth. It is displayed on the left axis of the figure as fraction. *Distance to parents* reports the straight-line distance in km between the individual's residence address and their parents' residence address. We use the minimum distance when parents have different residences. It is displayed on the right axis of the figure in km. All regressions are estimated using inverse probability weights to account for the over-representation of high-income families in urban areas. We define urban areas as the three largest Swedish municipalities (Stockholm, Gothenburg, and Malmö). Appendix Figure C10 displays results of the unweighted regressions. The estimation sample is based on 1,589,961 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. More details are provided in Section 2.

4.2 Mechanisms

Knowing that disadvantaged individuals face barriers across both time and space, we now examine the mechanisms through which these disadvantages materialize. We proceed systematically. First, we investigate human capital inequalities by asking whether differences in education or skills account for the observed disparities between groups. If not, a residual remains—implying that, conditional on education and skills, the groups experience different returns or outcomes.

Next, we turn to the labor market. We assess whether the groups differ in access to firms and occupations, or whether, within the same firms and occupations, they receive different earnings. Finally, we integrate the human capital and labor market dimensions by examining the residual variation in earnings, conditional on skills, education, firms, and occupations. Naturally, even this final residual may reflect unobserved skills correlated with family background.

Our empirical approach relies on estimating a series of Mincer equations to harmonize the units of the different measures. We then decompose total earnings into components explained by each factor, using the predicted values. We begin with the human capital exercise to illustrate this framework in practice.

4.2.1 Human Capital Inequality: Longer Education, Better Human Capital

To measure the contribution of education and skills, we use four measures of human capital. Utilizing a rich Swedish dataset on education *degree* and *field*, we decompose the differences in the quantity and quality of education. Regarding the degree, we exploit 13 distinct categories of highest completed education, ranging from pre-school level to doctoral degree (Appendix Table A.2). Regarding the field of study, we observe 336 categories, e.g., "Teacher training, industry and crafts" or "Industrial and product design" (Appendix Table A.1).

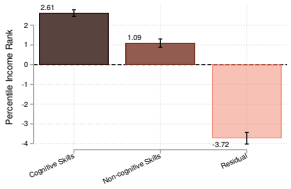
Moreover, we use the skill measures from Swedish draft data. For men, *cognitive* and *non-cognitive skills* are measured at age 18 as part of the military draft (Edin et al., 2022; Almgren et al., 2023; Hermo et al., 2022). We use four measures of cognitive skills and four measures of non-cognitive skills. Cognitive skill measures are based on test scores for (i) inductive skills, (ii) verbal comprehension, (iii) spatial ability, and (iv) technical understanding. Non-cognitive skill measures stem from an interview with a certified psychologist. They include scores for (i) emotional stability, (ii) intensity, (iii) psychological energy, and (iv) social maturity. These skill measures are available only for participants in military tests, i.e., men. Hence, we perform the regression only for men. We perform the equivalent exercise when omitting skill measures for women in Appendix Figure C5.

The challenge of using these four measures of human capital lies in their different measurement units. To address this, we standardize them by converting them into income-equivalent units. This involves regressing an individual's lifetime income rank on these measured facets of human capital using a non-parametric approach with fixed effects for each measure. The regression has an R^2 of 0.247 based on 560,167 observations. The predicted income ranks from the different measures of human capital serve as respective outcome variables of our main specification in Regression 11. Figure 8 illustrates these findings.

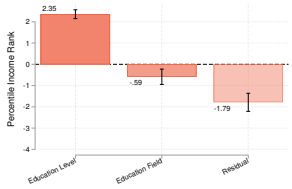
To reach the top decile, self-made individuals from bottom-decile backgrounds exhibit higher skills, encompassing both cognitive and non-cognitive aspects, by age 18. They opt for fields of study with lower average earnings but engage in longer periods of education. The beneficial impact of their extended education is roughly equivalent to the diminished earnings associated with their chosen fields. Thus, self-made individuals

offset the lower financial returns of their fields by pursuing lengthier educational paths. Consequently, when accounting solely for educational outcomes, all achievers are predicted to have similar incomes. The distinguishing factor, then, lies in their skill levels. The superior cognitive and non-cognitive skills of self-made individuals compensate for either the unobserved skills of their more affluent peers or the advantages those peers have, such as social networks and parental know-how. These aspects outside of an individual's measured human capital are captured in the residual component.

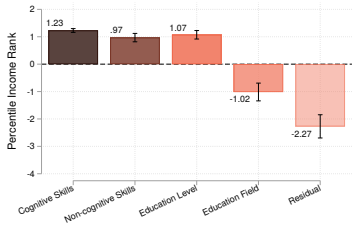
Figure 8: Mechanism: Divergence in educational paths



(a) Skill Inequality



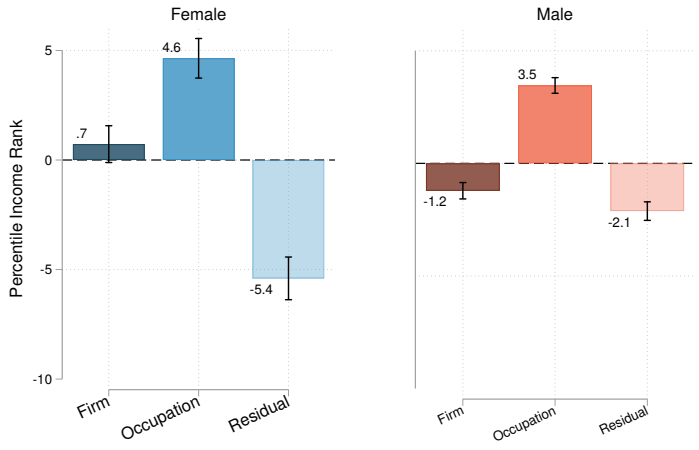
(b) Education Inequality



(c) Skill and Education Inequality

Note: This figure illustrates the relative differences in human capital between individuals reaching the top decile of lifetime income from different parental income deciles. It plots the coefficients $\{\beta_1\}$ from Regression 11, where the outcomes are four measures of human capital: cognitive skills, noncognitive skills, education level, and education field. Prior to using these measures as outcomes in Regression 11, we harmonize them to a common scale by mapping each to predicted lifetime-income ranks. Specifically, we estimate three separate Mincerian regressions of lifetime income rank: (i) on cognitive and noncognitive skills (Panel 8a), (ii) on education level and field (Panel 8b), and (iii) on the combination of both sets of variables (Panel 8c). This procedure decomposes lifetime income into human capital components. In Appendix Figure **??**, we add labor market outcomes (occupation and firm) to this decomposition. We restrict this exercise to men because it includes skill measures from the military draft (560,167 individuals). We report the corresponding estimates for females, without the skill measures, in Appendix Figures C5a and C5b.

Figure 9: Mechanism: Divergence in the labor market



(a) Labor Market Inequality

Note: This figure illustrates the relative difference in labor market outcomes between individuals reaching the top decile of lifetime income from different parental income deciles. It plots the coefficients $\{\beta_1\}$ from Regression 11, where the outcomes represent two aspects of the labor market: return to the individual's workplace (firm) and occupation. We focus on the main occupation and firm of the individual, defined as the mode of observed firms and occupations. Prior to using these measures as outcomes in Regression 11, we harmonize them to a common scale by mapping each to predicted lifetime income ranks. Specifically, we estimate a Mincerian regression of lifetime income rank on firm and occupation fixed effects. We use 123 occupation categories. In Appendix Figure **??**, we add human capital measures to this decomposition.

4.2.2 Labor Market Inequality: The Role of Firms and Occupations

We harmonize labor market measures using the same approach as before. Specifically, we estimate a Mincerian regression of lifetime income on firm and occupation fixed effects. Occupations are categorized into 123 groups, as detailed in Table A.3. To render labor market contributions comparable to those of human capital, we restrict the analysis to men. The male sample includes 42,180 distinct firms.[23] Corresponding estimates for women are reported in Figure C5b. The regression yields an R-Squared of 0.499.

Figure 9 depicts the role of labor market components for men from disadvantaged backgrounds. To reach the top decile of the income distribution, individuals from the bottom decile select, on average, more into high-return occupations. This positive occupational contribution to lifetime earnings is partially offset by their concentration in firms that, on average, pay lower returns.

This pattern persists even after controlling for human capital measures. As shown in Appendix Figure C4, the dominant positive contributor is the combination of lucrative occupations and relatively higher cognitive and non-cognitive skills among disadvantaged achievers, compared to their high-SES counterparts. The largest negative contributor is the regression residual, which likely captures unobserved factors such as limited access

[23] Industry-level differences are absorbed by the firm fixed effects.

to networks or job-search know-how. We elaborate on these mechanisms in the following section.

Which occupations and education study fields facilitate upward mobility for disadvantaged individuals? Table 2 lists the top five overrepresented occupations and study fields among achievers from top versus bottom decile parental backgrounds. Notably, the occupational and educational paths differ. High-SES individuals disproportionately enter professions such as medicine and law, whereas low-SES achievers attain high earnings in careers such as becoming a pilot or joining the military. Similarly, low-SES achievers are more likely to study mathematics or attain a military education as opposed to high-SES achievers who study medicine.

We further address the time and pleasure spent on work and leisure using survey data on the working environment stemming from the Swedish Work Environment Survey (AMU). We do this by measuring multiple dimensions of occupational amenities, including stress, work-family conflict, sleep disruptions, health, and job satisfaction. We construct occupation-level amenity measures by regressing each survey outcome on three-digit occupation fixed effects, with controls for education level, field of study, year of birth, and calendar-year-by-gender interactions. The estimated occupation fixed effects then measure the non-pecuniary characteristic of the occupation, net of observed worker skills.

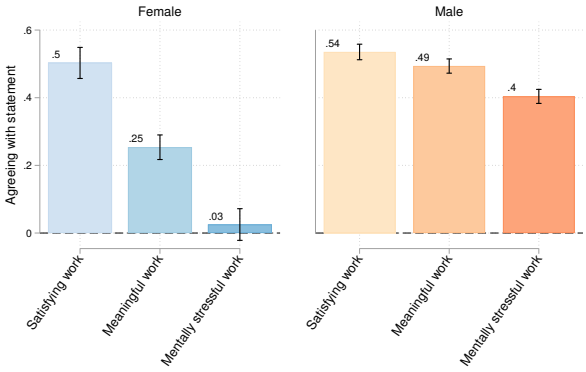
We find that disadvantaged achievers sort into occupations characterized by both positive and negative non-pecuniary attributes. Figure 10a documents that, relative to advantaged achievers, disadvantaged achievers are more likely to hold occupations associated with higher average levels of job satisfaction and perceived meaningfulness. At the same time, these occupations exhibit significantly higher levels of reported stress for men.

The elevated stress associated with these occupations extends beyond the workplace. Figure 10b examines occupation-level amenities that plausibly affect life outside of work, including the ability to relax during leisure time, time spent with family, sleep quality, and health (proxied by the frequency of headaches). Across all dimensions, occupations held by disadvantaged achievers score worse. These occupations are associated with higher probabilities of being unable to relax in free time, less time with family, poorer sleep patterns, and worse health outcomes.

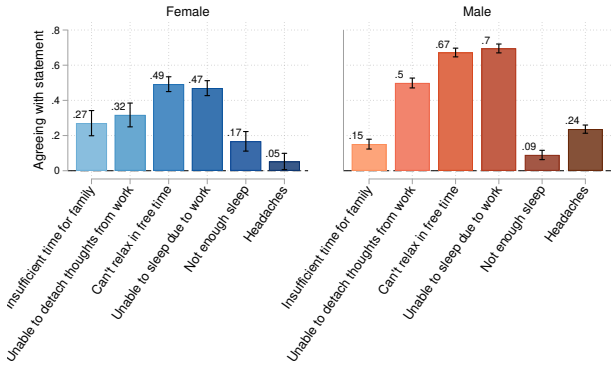
Taken together, the evidence indicates that disadvantaged achievers disproportionately enter occupations that, on average, are more rewarding along intrinsic dimensions such as satisfaction and meaningfulness, but that also impose greater stress and broader non-work costs. Similar income levels thus conceal differences in the non-pecuniary aspects of occupation, with disadvantaged achievers facing job rewards but also strains on

time, health, and family life.

Figure 10: Non-pecuniary Aspects of the Occupation



(a) Amenities associated with working life



(b) Amenities associated with private life

Note: This figure illustrates differences in occupation-level amenities and characteristics of the occupation. It plots the coefficient β_1 from Regression 11, where the dependent variables are alternative measures of occupational amenities. Occupational amenities are constructed from a biennial workplace survey (AMU) between 1997 and 2020, in which respondents rate statements about their work environment on a five-point scale. We aggregate individual-level survey responses to the three-digit occupation level by estimating regressions of each survey outcome on occupation fixed effects, controlling for indicators of education level, field of study, year of birth, and calendar year, all interacted with gender. We standardize each variable.

4.2.3 Steep Job Ladder

After accounting for firms, occupations, and different measures of human capital, residual disparities remain between privileged and selfmade achievers. This residual may reflect certain labor market know-how or access to networks that correlate with the background. We address this in Figure 11 which reports different dimensions of the working life. First, Figure 11a documents that selfmades start their career at later ages. To reach the top decile of the income distribution, those with parental income in the lowest decile are, on average, 0.4 to 0.8 years older at the time they start their first job than their high-

income peers from advantaged backgrounds.

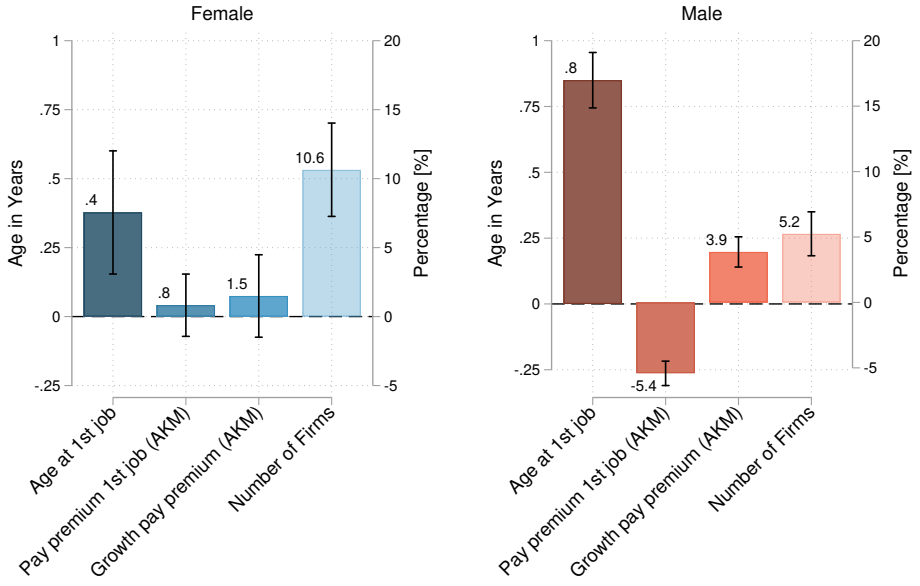
Furthermore, selfmades start their career in lower-paying firms, as judged by AKM firm fixed effects.[24] For male selfmade achievers, the first job has a 5 percent lower earnings premium compared with high-earners of highest-decile parents. However, self-mades change jobs more often, and their firm fixed effects grow more than their peers and with such a magnitude that by age 45 selfmades catch up with their peers. This is because the magnitude of the difference in their first firm is close to the magnitude of the growth.

Self-made achievers tend to possess higher levels of human capital, suggesting that privileged achievers may rely on additional mechanisms to enhance their career. One such mechanism is access to better social networks, which can assist them in identifying high-paying educational opportunities and firms.

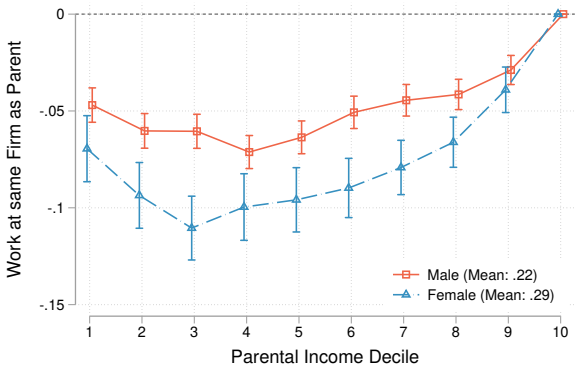
To investigate this hypothesis, we construct a proxy for family networks. Following the employer inheritance literature (see Corak and Piraino (2011)), Figure 11b illustrates that self-made achievers are six percentage points less likely to work for a firm any of tehir parents have worked for—a substantial effect, representing nearly 25 percent of the baseline likelihood. There is only limited evidence of reliance on classmate networks, which selfmades use to compensate for the lack of family networks (Appendix Figure 11c). This pattern supports the notion that privileged achievers can leverage additional resources beyond conventional human capital.

[24]We compute the AKM firm fixed effect by regressing log earnings on individual fixed effects and firm fixed effects, including controls of age and education, following Abowd et al. (1999); Card et al. (2013). We run this regression for the working population of Sweden using the largest connected set of firms. We interpret the firm fixed effect as the average firm pay premium.

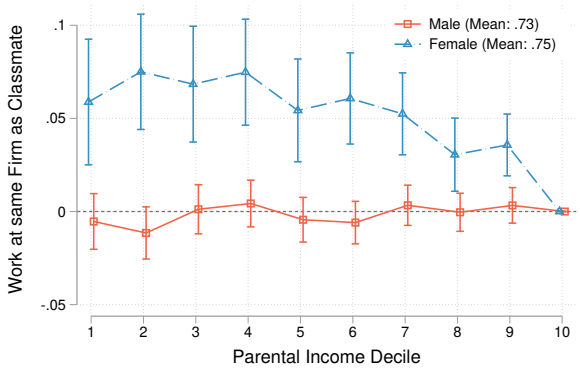
Figure 11: Underprivileged Achievers are Late Bloomers



(a) Differences in Career Trajectories



(b) Same Firm as Parent



(c) Same Firm as Classmate

Note: This figure illustrates the relative different career outcomes of individuals reaching the top-decile of lifetime income as a function of their parental income deciles ($\{\beta_l\}$ of Equation 11). Panel 11a reports $\{\beta_1\}$ for different type of career measures. *Firm Pay Premium* is measured as the AKM firm fixed effect in the fashion of Abowd et al. (1999); Card et al. (2013). *Number of firms* captures the total number of different firms between ages 30 and 45, due to the employer-employee dataset starting in 1985, when our oldest cohort was 30. The estimation sample is based on 1,514,601 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. The number is slightly below the main sample specification as we exclude individuals who have missing firm identifiers. Panel 11b and Panel 11c illustrate the differences in access to networks. We measure the outcome as having worked in the same firm between ages 30 and 45 as any firm their parents or their university peers have ever worked for. We define university peers as the group of individuals who graduate in the same year, the same university major, and the same municipality. The baseline mean reports the average of dependent variables among the bottom-decile-earning individuals with bottom-decile-earning parents. For more details, see Section 2.

5 Discussions

5.1 Quantity vs. Quality of Children

We saw that underprivileged achievers have, on average, fewer children. On the one hand, this could imply that they can spend more time per child, balancing the "quantity vs. quality" tradeoff. This would be in contrast to the evidence we saw on parental leave (Figures 5b). On the other hand, if underprivileged achievers spend more time building their careers, this could crowd out family time, affecting both the number of children and the time spent per child. Data sources to measure time spent with the family are limited. Hence, we proxy human capital investments into their children using achieved test scores in school. Specifically, we report the probability of the offspring passing all three subjects (English, Swedish, and Maths) in the national test exam in grade 9.

Figure C13 shows that the first child of underprivileged achievers is 4 percentage points more likely to pass all three subjects in the national exam compared to children from privileged achievers. This could either mean that underprivileged achievers invest more into the human capital of their (fewer) children (the channel we aim to test for) or that these children are more able.

To address the latter channel, we control in Figures C13b and C13c for parental skills, assuming that there is a correlation between parental and children's skills. The magnitude of the coefficient drops to 1 percentage points when controlling for parents' education level and field of study, and to 1 percentage point when further controlling for the father's cognitive and non-cognitive skills. Most coefficients turn statistically insignificant. This evidence suggests that the positive correlation is driven by the fact that children of selfmade achievers are, on average, more able. Conditional on parental skills, underprivileged achievers do not significantly invest more or less into their offspring.

5.2 Marriage & Assortative Mating

The delayed transition into family life, displayed in Figure 6, is accompanied by systematic differences in partner characteristics. Appendix Figure C7 shows that, conditional on reaching the top decile of the lifetime income distribution, individuals from lower-income parental backgrounds tend to match with spouses of higher observed socioeconomic status than do otherwise similar individuals from affluent families. The pattern is strongest for spouses' parental income rank and years of education, and is especially pronounced for women. At the bottom of the parental income distribution, the spouses of upwardly mobile women come from families with substantially higher parental income ranks than

the spouses of women from top-decile backgrounds. For both men and women, spouses are also roughly 0.5 to 1 year more educated. This must be compared with one's own higher education, Figures 6, which shows a similar magnitude of one additional year of education. By contrast, differences in spouses' own income rank are smaller and less precisely estimated. Overall, the figure points to substantial assortative matching among the upwardly mobile, especially along family background and education. Our interpretation is that finding a spouse with a higher background, but not necessarily higher education, is a way of climbing the ladder.

This pattern goes along with the results of high achievers spending less family time, as the extra time spent on building their career is not offset by the spouse spending more time at home.

5.2.1 Time Trend

We examine the change in the cost of mobility across 21 birth cohorts by dividing our sample into two periods: individuals born before 1965 and those born afterward (Figure C6). The number of children among underprivileged achievers is negative for both cohort groups; however, the severity decreases over time. Specifically, underprivileged achievers born between 1955 and 1965 have, on average, 0.15 fewer children, compared to 0.1 fewer children for those born between 1965 and 1975. While the difference is not statistically significant for the bottom decile, the change in magnitude may reflect shifts over this time period.

5.2.2 Beyond Family: Health Outcomes

We use family outcomes as our primary measure of the private life. Yet, there are other aspects of life not captured in this paper due to data limitations. This includes, for example, time spent in leisure or investments in health. We address the latter by leveraging Swedish register data on healthcare utilization. We measure health outcomes along two dimensions. First, we show in Figure C11 that we do not find a statistically significant correlation on mortality. Second, we estimate Equation 11 using the total days spent on sick leave as an outcome variable. Selfmade achievers have on average 1.5 fewer days of sick leave per year compared to their privileged counterparts (Figure C12a). To address the concern that we rather capture differences in healthcare utilization than actual differences in health between these two groups, we further report the within-person difference in sick leave days over the working life. Doing it this way, we do not find a statistical difference between the groups (Figure C12b). While this evidence suggests that health

outcomes of selfmade achievers are not worse than privileged achievers, we have to note that estimates are not precisely estimated meaning that we cannot conclusively rule out any effect on health.

6 Conclusion

In an early expression of the career–family tradeoff, in 1492, Cassandra Fedele—a prominent humanist writer—wrote to Alessandra Scala, a Florentine scholar: "And so, my Alessandra, you are uncertain whether to dedicate yourself to the Muses or to a Man?"[25] Both women managed to have both—a career and a family. They married and achieved enough success to be remembered, perhaps thanks to their privileged backgrounds, as argued in this paper.

This paper reexamines intergenerational mobility through the lens of welfare, rather than income alone, and shows that the cost of upward mobility is not equally borne. We develop a theoretical model that distinguishes between income and welfare mobility by incorporating a career-family tradeoff that varies with parental background. Using rich Swedish administrative data, we document that individuals from low-income families pay a higher price to climb the income ladder—having fewer children, delaying family formation, and spending less time with each child. These disparities stem from more demanding paths to success: longer studies, more frequent geographic moves, and fewer business connections. This suggests that disadvantaged individuals must exert greater effort and forgo more of their personal lives to achieve the same economic outcomes as achievers from high-income families.

[25] "Muses" refers to the Greek goddesses of inspiration in the arts, symbolizing creative or literary pursuits (see Kaborycha, 2016).

References

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

ADDA, J., C. DUSTMANN, AND K. STEVENS (2017): "The career costs of children," *Journal of Political Economy*, 125, 293–337.

ALMGREN, M., J. KRAMER, AND J. SIGURDSSON (2023): "It Runs in the Family: Occupational Choice and the Allocation of Talent," .

ALTHOFF, L., H. B. GRAY, AND H. REICHARDT (2025): "Mothers, Schools, and the Making of American Human Capital Mobility," .

ANGELOV, N., P. JOHANSSON, AND E. LINDAHL (2016): "Parenthood and the gender gap in pay," *Journal of labor economics*, 34, 545–579.

ATKINSON, A. B. (1980): "Intergenerational income mobility," *JHS Journal*, 3, 61–73.

BEAUD, S., J. LINDGAARD, AND J. CONFAVREUX (2013): *La france invisible*, La découverte.

BECKER, G. S. AND B. R. CHISWICK (1966): "Education and the Distribution of Earnings," *The American Economic Review*, 56, 358–369.

BELL, A., R. CHETTY, X. JARAVEL, N. PETKOVA, AND J. VAN REENEN (2019): "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 134, 647–713.

BERGER, T., P. ENGZELL, B. ERIKSSON, AND J. MOLINDER (2023): "Social mobility in Sweden before the welfare state," *The Journal of Economic History*, 83, 431–463.

BJÖRKLUND, A. AND M. JÄNTTI (1997): "Intergenerational income mobility in Sweden compared to the United States," *The American Economic Review*, 87, 1009–1018.

BJÖRKLUND, A., M. JÄNTTI, ET AL. (2009): "Intergenerational income mobility and the role of family background," in *Oxford Handbook of Economic Inequality*, Oxford University Press, vol. 491, 521.

BLACK, S. E. AND P. J. DEVEREUX (2011): "Recent developments in intergenerational mobility," *Handbook of labor economics*, 4, 1487–1541.

BOAR, C. AND D. LASHKARI (2021): "Occupational choice and the intergenerational mobility of welfare," Tech. rep., National Bureau of Economic Research.

BOLTON, P. AND M. DEWATRIPONT (2005): *Contract Theory*, Cambridge, MA: MIT Press.

CARD, D., J. HEINING, AND P. KLINE (2013): "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly journal of economics*, 128, 967–1015.

CASARICO, A. AND S. LATTANZIO (2023): "Behind the child penalty: understanding what contributes to the labour market costs of motherhood," *Journal of Population Economics*, 36, 1489–1511.

CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014): "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 129, 1553–1623.

CORAK, M. AND P. PIRAINO (2011): "The Intergenerational Transmission of Employers," *Journal of Labor Economics*, 29, 37–68.

D'AGOSTINO, M. AND V. DARDANONI (2009): "The measurement of rank mobility," *Journal of Economic Theory*, 144, 1783–1803.

DOBBIN, C. AND T. ZOHAR (2023): "Quantifying the role of firms in intergenerational mobility," *Unpublished manuscript*.

EDIN, P.-A., P. FREDRIKSSON, M. NYBOM, AND B. ÖCKERT (2022): "The rising return to noncognitive skill," *American Economic Journal: Applied Economics*, 14, 78–100.

EIKA, L., M. MOGSTAD, AND B. ZAFAR (2019): "Educational assortative mating and household income inequality," *Journal of Political Economy*, 127, 2795–2835.

ENGZELL, P. AND N. WILMERS (2021): "Firms and the intergenerational transmission of labor market advantage," *Preprint, SocArXiv December*.

GOLDIN, C. (1997): "Career and Family: College Women Look to the Past," *Gender and Family Issues in the Workplace*.

——— (2014): "A grand gender convergence: Its last chapter," *American economic review*, 104, 1091–1119.

——— (2021): *Career and family: Women's century-long journey toward equity*, Princeton University Press.

HARRINGTON, E. AND M. E. KAHN (2023): "Has the rise of work-from-home reduced the motherhood penalty in the labor market," Tech. rep., Technical report.

HARRINGTON, M. (1997): *The other america*, Simon and Schuster.

HERMO, S., M. PÄÄLLYSAHO, D. SEIM, AND J. M. SHAPIRO (2022): "Labor market returns and the evolution of cognitive skills: Theory and evidence," *The Quarterly Journal of Economics*, 137, 2309–2361.

HOTZ, V. J., P. JOHANSSON, AND A. KARIMI (2018): "Parenthood, family friendly workplaces, and the gender gaps in early work careers," Tech. rep., National Bureau of Economic Research.

JÄNTTI, M., B. BRATSBERG, K. ROED, O. RAAUM, R. NAYLOR, E. OSTERBACKA, A. BJORKLUND, AND T. ERIKSSON (2006): "American exceptionalism in a new light: a comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States," .

JONES, C. I. (2022): "The past and future of economic growth: A semi-endogenous perspective," *Annual Review of Economics*, 14, 125–152.

KABORYCHA, L., ed. (2016): *Corresponding Renaissance: Letters Written by Italian Women, 1375-1650*, Oxford: Oxford University Press, edited and translated by Lisa Kaborycha.

KLEVEN, H., C. LANDAIS, J. POSCH, A. STEINHAUER, AND J. ZWEIMÜLLER (2019a): "Child penalties across countries: Evidence and explanations," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 109, 122–126.

KLEVEN, H., C. LANDAIS, AND J. E. SØGAARD (2019b): "Children and gender inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 11, 181–209.

KOCHAR, S. C. (2022): *Stochastic Comparisons with Applications: In Order Statistics and Spacings*, Springer Nature.

LAFFONT, J.-J. AND D. MARTIMORT (2002): *The Theory of Incentives: The Principal–Agent Model*, Princeton, NJ: Princeton University Press.

LONG, J. AND J. FERRIE (2013): "Intergenerational occupational mobility in Great Britain and the United States since 1850," *American Economic Review*, 103, 1109–1137.

MAJOR, L. E. AND S. MACHIN (2018): *Social mobility: And its enemies*, Penguin UK.

MAS-COLELL, A. (1985): *The Theory of General Economic Equilibrium: A Differentiable Approach*, Econometric Society Monographs, Cambridge University Press.

MILGROM, P. AND J. SHANNON (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.

MOGSTAD, M. AND G. TORSVIK (2023): "Family background, neighborhoods, and intergenerational mobility," *Handbook of the Economics of the Family*, 1, 327–387.

NAKAMURA, E., J. SIGURDSSON, AND J. STEINSSON (2022): "The gift of moving: Intergenerational consequences of a mobility shock," *The Review of Economic Studies*, 89, 1557–1592.

NELSEN, R. B. (2006): *An introduction to copulas*, Springer.

NELSEN, R. B., J. J. QUESADA-MOLINA, J. A. RODRÍGUEZ-LALLENA, AND M. ÚBEDA FLORES (2003): "Kendall Distribution Functions," *Statistics & Probability Letters*, 65, 263–268.

NICO, M. (2021): "Identity and change of a field: A literature analysis of the concept of social mobility," *Social Science Information*, 60, 457–478.

PETERS, H. E. (1992): "Patterns of intergenerational mobility in income and earnings," *The review of Economics and Statistics*, 456–466.

PUTNAM, R. D. (2016): *Our kids: The American dream in crisis*, Simon and Schuster.

SALANIÉ, B. (2005): *The Economics of Contracts: A Primer*, Cambridge, MA: MIT Press, 2 ed.

SCHWEIZER, B. AND E. F. WOLFF (1981): "On nonparametric measures of dependence for random variables," *The annals of statistics*, 9, 879–885.

SOLON, G. (1999): "Intergenerational mobility in the labor market," in *Handbook of labor economics*, Elsevier, vol. 3, 1761–1800.

SOROKIN, P. A. (1927): *Social and Cultural Mobility*, The Free Press of Glencoe, Illinois.

VAN LEEUWEN, M. H. AND I. MAAS (2010): "Historical studies of social mobility and stratification," *Annual Review of Sociology*, 36, 429–451.

Table 1: Sample Characteristics

	Mean	SD	Min	Max
Female	0.48	(0.50)	0.00	1.00
Year of Birth	1965.14	(6.01)	1955.00	1975.00
Number of siblings	1.90	(1.25)	0.00	7.00
Education >= college	0.34	(0.47)	0.00	1.00
Completed Fertility	1.90	(1.12)	0.00	5.00
Any Children	0.85	(0.36)	0.00	1.00
Days of parental leave per child	213.24	(182.12)	0.00	699.00
Age at first child	28.67	(5.48)	17.00	44.00
Years of Education	12.51	(2.26)	7.00	20.00
Age at first Employment	26.02	(5.38)	20.00	65.00
Age marriage | married before	31.03	(5.68)	18.00	45.00
Rank of Birthplace Municipality	0.50	(0.06)	0.27	0.63
Moved from Birthplace	0.31	(0.46)	0.00	1.00
Number of Moves between age 35-45	0.89	(1.17)	0.00	10.00
Distance to Parent at Age 35 (km)	67.23	(140.90)	0.00	1442.91
Number of Firms	3.16	(1.90)	0.00	16.00
Work at same Firm as Parent between Age 30-45	0.22	(0.42)	0.00	1.00
Work at same Firm as Classmate between Age 30-45	0.72	(0.45)	0.00	1.00
Observations	1589961			

Note: The table reports sample characteristics of the main estimation sample. The estimation sample is 1,589,961 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden.

Table 2: Occupations and Study Fields Among Underprivileged Achievers

Top Decile Parents	Overrep.	**Bottom Decile Parents**	Overrep.
Panel A: Occupations			
Chief Executives	10.02	Senior Officials/Politicians	6.67
Health/Medical Specialists	8.81	Pilots	5.91
Lawyers	7.04	Specialized Managers	5.69
Specialized Managers	6.91	Chief Executives	5.29
Senior Officials/Politicians	6.64	Military	4.44
Panel B: Study Fields			
Medicine	8.93	Mathematics	5.34
Ind. Engineering & Organization	8.16	Military Education	4.66
Materials and Manufacturing	8.06	Statistics	4.21
Engineering & Technical Industry	6.62	Engineering & Technical Industry	3.82
Economics and Economic History	5.43	Ind. Engineering & Organization	3.66

Note: This table reports the top five overrepresented occupations (Panel A) and study fields (Panel B) for advantaged (disadvantaged) achievers in the left (right) columns. Overrepresentation is defined as the ratio of a category's share within the group to its share in the full sample.

Online Appendix of:

Who Can Have It All?

The Unequal Cost of Career & Inter-generational Mobility

Patrizia Massner & Arash Nekoei

March 27, 2026

Appendix contents

A Theory **A3**

A.1 Definition of IGM: Discordance and Kendall's Rank Correlation A3

A.2 Detailed Setting A4

A.2.1 Assumption for the model A4

A.2.2 DGP of History and Talent A5

A.2.3 Talent Metric A6

A.3 Proofs (except the Theorem) A7

A.3.1 Claim 1 . . . . . A7

A.3.2 Appendix Claim A1: Index Representation A7

A.3.3 Claim 2 . . . . . A8

A.3.4 Proposition 1 . . . . . A10

A.3.5 Corollary 2 . . . . . A12

A.3.6 Other claims A13

A.4 Theorem 1 . . . . . A18

A.4.1 Two Lemmas A18

A.4.2 The Proof of Theorem 1 . . . . . A22

A.5 Case Studies A26

A.5.1 Time-Invariant Background: Binary Case A26

A.5.2 Parental Career as background $h = z_{-1}$ A30

A.5.3 Markovian joint log-normal DGP for talent A34

B Models and Codes **A38**

B.1 Model 2 . . . . . A38

B.2 Model 3 . . . . . A39

C Appendix Figures **A41**

D Appendix Tables **A54**

A Theory

We denote with capital letters the random variables (r.v.) and use $\widetilde{X}$ for an independent copy of a r.v. X.

A.1 Definition of IGM: Discordance and Kendall's Rank Correlation

We measure intergenerational mobility (IGM) as the probability that two dynasties reverse their relative ranks across generations. Statistically, this is the probability of *discordance*, closely linked to Kendall's rank correlation coefficient τ. This section introduces the concept of discordance and explains its relationship to Kendall's τ.

Let (X, Y) be a pair of random variables and $(\widetilde{X}, \widetilde{Y})$ an independent copy. The pairs are *concordant* if the two observations move in the same direction, that is

$$(\widetilde{X} - X)(\widetilde{Y} - Y) > 0,$$

and *discordant* if the two observations move in the opposite direction, that is

$$(\widetilde{X} - X)(\widetilde{Y} - Y) < 0.$$

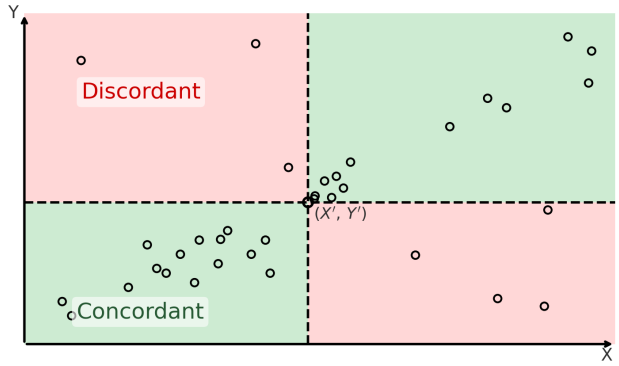


Figure A1: Discordant vs. Concordant

Assume (X, Y) has a continuous joint distribution (no ties), so that $\mathbb{P}(\text{discordant}) + \mathbb{P}(\text{concordant}) = 1$. By exchangeability of (X, Y) and $(\widetilde{X}, \widetilde{Y})$,

$$\mathbb{P}(\text{discordant}) = \mathbb{P}((\widetilde{X} - X)(\widetilde{Y} - Y) < 0) \quad = 2\,\mathbb{P}(X > \widetilde{X},\, Y < \widetilde{Y}) = \mathbb{P}(X > \widetilde{X} \mid Y < \widetilde{Y}),$$

since $\mathbb{P}(Y < \widetilde{Y}) = \frac{1}{2}$ under continuity. Kendall's τ is defined as the difference between the probabilities of concordance and discordance,

$$\tau := \mathbb{P}\big((\widetilde{X} - X)(\widetilde{Y} - Y) > 0\big) - \mathbb{P}\big((\widetilde{X} - X)(\widetilde{Y} - Y) < 0\big) = \mathbb{E}\Big[\operatorname{sign}\big((\widetilde{X} - X)(\widetilde{Y} - Y)\big)\Big],$$

hence

$$\tau = 1 - 2\,\mathbb{P}(\text{discordant}) = 2\,\mathbb{P}(\text{concordant}) - 1,$$

and, again by symmetry,

$$\tau = 1 - 4\,\mathbb{P}(X > \widetilde{X},\, Y < \widetilde{Y}) = 4\,\mathbb{P}(X < \widetilde{X},\, Y < \widetilde{Y}) - 1.$$

Let $F_{X,Y}$ denote the joint CDF of (X, Y). Then

$$\mathbb{P}(X < \widetilde{X},\, Y < \widetilde{Y}) = \iint F_{X,Y}(x, y)\, dF_{X,Y}(x, y) = \mathbb{E}\big[F_{X,Y}(X, Y)\big],$$

Hence, the Kendall Tau can be written as $\tau = 4\,\mathbb{E}\big[F_{X,Y}(X, Y)\big] - 1$.[26] Denoting by $C_{X,Y}$ the copula of (X, Y), we can write the Kendall Tau as

$$\tau = 4\,\mathbb{E}\Big[C_{X,Y}\big(F_X(X), F_Y(Y)\big)\Big] - 1,$$

given that $F_{X,Y}(x, y) = C_{X,Y}(F_X(x), F_Y(y))$. Consequently,

$$\mathbb{P}(\text{discordant}) = \frac{1 - \tau}{2} = 1 - 2\,\mathbb{E}\Big[C_{X,Y}\big(F_X(X), F_Y(Y)\big)\Big]. \tag{A1}$$

This representation shows that our IGM measure depends only on the copula and is therefore invariant to strictly increasing marginal transformations of X and Y.

We end this section with a note on the Gaussian case. For a bivariate normal vector (X, Y) with Pearson correlation ρ, the three standard rank–based and linear correlations are in one–to–one correspondence. In particular, Kendall's τ and Spearman's ρ_S are explicit functions of ρ:

$$\tau = \frac{2}{\pi}\arcsin(\rho), \qquad \rho_S = \frac{6}{\pi}\arcsin\Big(\frac{\rho}{2}\Big). \tag{A2}$$

A.2 Detailed Setting

A.2.1 Assumption for the model

Assumption A1. *Assume the utility function $u : \mathbb{R}_+^2 \to \mathbb{R}$ is continuous, differentiable, componentwise strictly increasing, and strictly concave; both production functions $F_z, F_k : \mathbb{R}_+^L \times \Theta \to \mathbb{R}_+$ continuous in (x, θ), differentiable, concave and componentwise increasing in their first argument (inputs); for each θ, the composite objective*

$$(x, y) \mapsto u\big(F_z(x, \theta), F_k(y, \theta)\big)$$

[26] "Kendall distribution function" K is defined as the distribution function of $F_{X,Y}(X, Y)$. Then $\tau = 3 - 4\int_0^1 K(t)\,dt$ and $K(t)$ is used to Kendall order, See Nelsen et al. (2003).

is strictly concave on $\mathbb{R}_+^{2L}$.

The endowment $T(\theta) \in \mathbb{R}_+^L$ has finite components, so the feasible set $(x, y) \in \mathbb{R}_+^{2L} : x + y \leq T(\theta)$ is nonempty, closed, and bounded. The objective is continuous and strictly concave on a compact convex set, implying that (1) has a unique maximizer.

Assumption A2. *Endow* $\mathbb{R}_+^L$ *with its Borel* σ*-algebra. Assume* θ *living in measurable space* $(\Theta, \mathcal{F})$, $T : \Theta \to \mathbb{R}_+^L$ *is* $\mathcal{F}$*-measurable and* F_z *and* F_k *both* $: \mathbb{R}_+^L \times \Theta \to \mathbb{R}_+$ *are* $\mathcal{F} \otimes B(\mathbb{R}_+^L)$*-measurable and continuous in inputs.*

Then the feasible correspondence $\Gamma(\theta) = \{(x, y) \in \mathbb{R}_+^{2L} : x + y \leq T(\theta)\}$ is measurable with nonempty compact values, because

$$\{(\theta, x, y) : (x, y) \in \Gamma(\theta)\} = \{(\theta, x, y) : x + y - T(\theta) \leq 0\}$$

is a measurable subset of $\Theta \times \mathbb{R}_+^{2L}$.

So by the measurable maximum theorem the value function $v(\theta)$ is $\mathcal{F}$-measurable and, the optimizer $(x(\theta), y(\theta))$ is $\mathcal{F}$-measurable. Consequently, $z(\theta) = F_z(x(\theta), \theta)$ and $k(\theta) = F_k(y(\theta), \theta)$ are $\mathcal{F}$-measurable as well.

A.2.2 DGP of History and Talent

We work on an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ rich enough to support all random objects introduced below (possibly after passing to a product extension, which we suppress in notation). All random variables and processes are defined on this space unless stated otherwise.

1. History. Let $H : \Omega \to \mathbb{R}$ denote (time-invariant) family background, and let $\widetilde{H}$ be an i.i.d. copy of H, independent of all other primitives. We identify "more advantaged background" with a larger real number.

2. Talent. Let $\{W_t\}_{t \in \mathbb{N}}$ be a strictly stationary, atomless process with support $\mathbb{R}_+ := [0, \infty)$, independent of history H. Let F be the marginal cdf of W_t, assumed strictly increasing on $\mathbb{R}_+$, with Lebesgue density $f \in C^\infty((0, \infty))$ and $\|f\|_\infty := \sup_{x \geq 0} f(x) < \infty$. Since W_t is atomless, $\mathbb{P}(W_t = 0) = 0$, and all density-based expressions are understood on $(0, \infty)$. Let $\{\widetilde{W}_t\}_{t \in \mathbb{Z}}$ be an i.i.d. copy of $\{W_t\}_{t \in \mathbb{Z}}$, independent of $\{W_t\}_{t \in \mathbb{Z}}$ and of $(H, \widetilde{H})$.

When only two consecutive dates are under consideration, we write $W := W_t$ and $W_{-1} := W_{t-1}$, and analogously for all derived variables (e.g. $U := U_t$, $Z := Z_t$, etc.). By strict stationarity, the law of (W_{t-1}, W_t) does not depend on t.

Define $U_t := F(W_t)$ and $\widetilde{U}_t := F(\widetilde{W}_t)$. Then $U_t, \widetilde{U}_t \sim \mathrm{Unif}[0, 1]$. By Sklar's theorem, there exists a unique copula $C : [0, 1]^2 \to [0, 1]$ such that for all $u_{-1}, u \in [0, 1]$,

$$\mathbb{P}(U_{-1} \leq u_{-1},\ U \leq u) = C(u_{-1}, u).$$

Assume $C \in C^2((0,1)^2)$ and denote partial derivatives by $\partial_i C$ and $\partial_{ij} C$. A regular conditional distribution kernel of U given $U_{-1} = u_{-1}$ is

$$\mathbb{P}(U \le u \mid U_{-1} = u_{-1}) = \partial_1 C(u_{-1}, u), \qquad u_{-1} \in (0,1),\ u \in [0,1],$$

with conditional density $c(u_{-1}, u)$ on $(0,1)$. Consequently, for $x > 0$ and $y \ge 0$,

$$\mathbb{P}(W \le y \mid W_{-1} = x) = \partial_1 C(F(x), F(y)), \qquad f_{W|W_{-1}=x}(y) = \partial_{12} C(F(x), F(y))\, f(y).$$

Whenever we write a conditional distribution given $U_{-1} = u_{-1}$ (or $W_{-1} = x$), we mean a fixed version of the regular conditional distribution. Under our smoothness assumptions on C, we may (and do) choose a version such that the identities below hold for Lebesgue-a.e. $u_{-1} \in (0,1)$ (equivalently, for F-a.e. $x \in (0,\infty)$). Since all subsequent results are unaffected by changes on null sets, we suppress the "a.e." qualifier.

A.2.3 Talent Metric

Let $z, v : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ be continuous and strictly increasing in w for each h. Define $Z_t := z(H, W_t)$ and $V_t := v(H, W_t)$, and likewise $\widetilde{Z}_t := z(\widetilde{H}, \widetilde{W}_t)$ and $\widetilde{V}_t := v(\widetilde{H}, \widetilde{W}_t)$. All are outcomes atomless.

For $y \in \{z, v\}$ and $(h, \tilde{h}, w) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$ define

$$\mu^y(\tilde{h}; h, w) := \inf\Big\{ w' \ge 0 :\ y(\tilde{h}, w') \ge y(h, w) \Big\}.$$

This is well-defined (finite) whenever $\lim_{w\to\infty} y(\tilde{h}, w) = +\infty$.

Assumption A3 (Common image). *For each $y \in \{z, v\}$, $y(h, 0) = 0$ for all $h \in \mathbb{R}$, and $\lim_{w\to\infty} y(h, w) = +\infty$ for all $h \in \mathbb{R}$. Equivalently, $\mathrm{Im}(y(h, \cdot)) = [0, \infty)$ for all h.*

The important part of this assumption is that the lowest talent receives the same regardless of their background, which is a strong assumption.

Under Assumption A3, $\mu^y(\tilde{h}; h, w)$ is the unique w' satisfying $y(h, w) = y(\tilde{h}, w')$, i.e.

$$\mu^y(\tilde{h}; h, w) = \big(y(\tilde{h}, \cdot)\big)^{-1}\big(y(h, w)\big).$$

Without Assumption A3, the definition implies $\mu^y(\tilde{h}; h, w) = 0$ whenever $y(h, w) \le y(\tilde{h}, 0)$ (the target outcome is below the minimum attainable under $\tilde{h}$).

If, in addition, $w \mapsto y(h, w)$ and $w \mapsto y(\tilde{h}, w)$ are C^2 on $(0, \infty)$ with $\partial_w y(\tilde{h}, w) > 0$, then for each fixed $(h, \tilde{h})$ the map $w \mapsto \mu^y(\tilde{h}; h, w)$ is C^2 on $\{w :\ y(h, w) > y(\tilde{h}, 0)\}$ (and on $(0, \infty)$ under Assumption A3).

Without Assumption above, the talent metric equal to zero when the output is too low to be attainable, which occurs when $h < \widetilde{h}$ and w is low. We prefer the common image assumption as

the generalized talent metric is not strictly increasing, which complicates the proof unnecessarily. Another solution has been to assume that the domain of W is the entire R. [Does this create a problem in the other part of the paper?]

A.3 Proofs (except the Theorem)

A.3.1 Claim 1

Fix $h' \succ h$ and w. First note that $\mu^y(h; h, w) = w$ and $\mu^y\left(\widetilde{h}; h, w\right) > w$ for any w and $y \in z, k, v$. For any w', strict monotonicity of $k(h', \cdot)$ implies

$$k(h, w) \geq k(h', w') \iff k(h', \mu^k(h'; h, w)) \geq k(h', w') \iff \mu^k(h'; h, w) \geq w',$$

and strict monotonicity of $z(h', \cdot)$ implies

$$z(h, w) > z(h', w') \iff z(h', \mu^z(h'; h, w)) > z(h', w') \iff \mu^z(h'; h, w) > w'.$$

Therefore, Assumption 2 is equivalent to

$$\mu^k(h'; h, w) \geq w' \Rightarrow \mu^z(h'; h, w) > w',$$

for all w'. This holds if and only if $\mu^z(h'; h, w) > \mu^k(h'; h, w)$: the "if" direction is immediate, and the "only if" direction follows by setting $w' = \mu^k(h'; h, w)$. The case $h \succ h'$ is symmetric and yields $\mu^k(h'; h, w) > \mu^z(h'; h, w)$.

A.3.2 Appendix Claim A1: Index Representation

Claim A1 (Index representation and dimensional reduction)**.** *Under Assumption 3, there exists a C^2 function $\beta : \mathcal{H} \to \mathbb{R}$ such that, for all $h, h' \in \mathcal{H}$, $h' \succeq h \iff \beta(h') \geq \beta(h)$. Moreover, for each outcome $y \in \{z, k, v\}$ there exists a function $\hat{y} : \beta(\mathcal{H}) \times \mathbb{R}_+ \to \mathbb{R}$ such that $y(h, w) = \hat{y}(\beta(h), w)$ for all $(h, w) \in \mathcal{H} \times \mathbb{R}_+$.*

Proof. By Assumption 3(i) and Proposition 2.3.9 in Mas-Colell (1985), there exists a C^2 utility index $\beta : \mathcal{H} \to \mathbb{R}$ representing $\succeq$, i.e., for all $h, h' \in \mathcal{H}$, $h' \succeq h \iff \beta(h') \geq \beta(h)$.

Fix $y \in \{z, k\}$. For each $b \in \beta(\mathcal{H})$ and $w \in \mathbb{R}_+$, choose any $h \in \mathcal{H}$ with $\beta(h) = b$ and define $\hat{y}(b, w) := y(h, w)$. $\hat{y}$ is well defined: let $h' \in \mathcal{H}$ be another history with $\beta(h') = b$. Then $\beta(h') = \beta(h)$ implies $h' \sim h$. By Assumption 3(ii), $y(h, w) = y(h', w)$ for all $w \in \mathbb{R}_+$. Hence $\hat{y}(b, w)$ does not depend on the particular choice of h with $\beta(h) = b$, and $\hat{y}$ is well defined. By construction, for any $(h, w) \in \mathcal{H} \times \mathbb{R}_+$, $y(h, w) = \hat{y}(\beta(h), w)$, which proves the claimed dimensional reduction. $Q.E.D.$

Remark 1 (Working with the reduced-form representation). *Let $B \coloneqq \beta(\mathcal{H}) \subset \mathbb{R}$. Under Assumption 3 and Lemma A1, each outcome $y \in \{z, k, v\}$ admits a scalar-index representation*

$$y(h, w) = \hat{y}(\beta(h), w) = \hat{y}(b, w) \quad \textit{with } b = \beta(h).$$

Thus, without loss of generality, we can work in the reduced space of pairs $(b, w) \in B \times \mathbb{R}_+$, with b summarizing the dynastic history h via the index β.

[connect to Assumption A3]

A.3.3 Claim 2

Proof. **First Proof of Claim 2**
Under Assumption 3, the index function β exists and $\hat{f}$ for $f \in \{c, k, v\}$ is defined on $\mathbb{R}^2$. Define $\hat{\varphi}$ as $\hat{\varphi}(\beta(h), w, t) = \varphi(h, w, t)$. $\hat{\varphi}$ is completely regular and twice continuously differentiable with $\frac{\partial}{\partial w}\hat{\varphi} > 0$. So We appeal to Theorem 3 in Milgrom and Shannon (1994). It implies the equivalence of Milgrom-Shannon's and Spence–Mirrlees' single crossing conditions of $\hat{\varphi}$. Moreover, we have shown that Assumption 2 is equivalent to Milgrom-Shannon's single crossing of φ and Assumption 4 is equivalent to Spence–Mirrlees' single crossing of φ. The last piece of the argument is that the single crossing of φ and $\hat{\varphi}$ is equivalent under Assumption 3. $Q.E.D.$

Proof. **Second Proof of Claim 2**

Assumption 2 $\Rightarrow$ Assumption 4
Define

$$\mu^{\Delta}(\widetilde{b}; b, w) := \mu^{z}(\widetilde{b}, b, w) - \mu^{k}(\widetilde{b}, b, w). \tag{A3}$$

Hence $\mu^{\Delta}(b; b, w) = 0$. By Assumption 2,

$$(\widetilde{b} - b)\mu^{\Delta}(\widetilde{b}; b, w) > 0, \tag{A4}$$

if $\widetilde{b} \neq b$. Therefore, for every $\varepsilon > 0$,

$$\frac{\mu^{\Delta}(b + \varepsilon; b, w) - \mu^{\Delta}(b; b, w)}{\varepsilon} = \frac{\mu^{\Delta}(b + \varepsilon; b, w)}{\varepsilon} \geq 0,$$

and also

$$\frac{\mu^{\Delta}(b - \varepsilon; b, w) - \mu^{\Delta}(b; b, w)}{-\varepsilon} = \frac{\mu^{\Delta}(b - \varepsilon; b, w)}{-\varepsilon} \geq 0.$$

Because μ^{Δ} is smooth in $\widetilde{b}$, its derivative exists at $\widetilde{b} = b$, and the one-sided inequalities imply $\partial_{\widetilde{b}}\mu^{\Delta}(b; b, w) \geq 0$.

Next, for $y \in \{z, k\}$, differentiate the identity $y\big(\widetilde{b}, \mu^y(\widetilde{b}, b, w)\big) = y(b, w)$ with respect to $\widetilde{b}$:

$$\partial_{\widetilde{b}} \mu^y(\widetilde{b}, b, w) = -\frac{\partial_b y}{\partial_w y}\big(\widetilde{b}, \mu^y(\widetilde{b}, b, w)\big).$$

Evaluating at $\widetilde{b} = b$ and using $\mu^y(b, b, w) = w$ gives

$$\partial_{\widetilde{b}} \mu^y(b, b, w) = -\frac{\partial_b y(b, w)}{\partial_w y(b, w)}.$$

Therefore

$$\partial_{\widetilde{b}} \mu^{\Delta}(b; b, w) = \partial_{\widetilde{b}} \mu^z(b, b, w) - \partial_{\widetilde{b}} \mu^k(b, b, w) = -\frac{\partial_b z(b, w)}{\partial_w z(b, w)} + \frac{\partial_b k(b, w)}{\partial_w k(b, w)}.$$

Since $\partial_{\widetilde{b}} \mu^{\Delta}(b; b, w) \geq 0$, we prove Assimption 4.

Assumption 2 $\Leftarrow$ Assumption 4

Fix (b, w) and define

$$m^z(\tau) := \mu^z(\tau, b, w), \qquad m^k(\tau) := \mu^k(\tau, b, w).$$

Now consider the auxiliary function

$$\Psi(\tau) := k\big(\tau, m^z(\tau)\big).$$

Differentiating,

$$\Psi'(\tau) = \partial_b k\big(\tau, m^z(\tau)\big) + \partial_w k\big(\tau, m^z(\tau)\big)\,(m^z)'(\tau).$$

From the identity $z\big(\tau, m^z(\tau)\big) = z(b, w)$, we obtain by implicit differentiation

$$(m^z)'(\tau) = -\frac{\partial_b z}{\partial_w z}\big(\tau, m^z(\tau)\big).$$

Substituting into the expression for $\Psi'(\tau)$ gives

$$\Psi'(\tau) = \partial_w k\big(\tau, m^z(\tau)\big)\left[\frac{\partial_b k}{\partial_w k} - \frac{\partial_b z}{\partial_w z}\right]\big(\tau, m^z(\tau)\big).$$

By assumption, the bracket is positive at every point, and $\partial_w k > 0$, so $\Psi'(\tau) > 0$for all τ. Hence Ψ is increasing.

Also, $m^z(b) = w$ and $m^k(b) = w$, so $\Psi(b) = k\big(b, m^z(b)\big) = k(b, w)$. On the other hand, by definition of m^k,

$$k\big(\tau, m^k(\tau)\big) = k(b, w) = \Psi(b) \qquad \text{for all } \tau.$$

Now let $\widetilde{b} > b$. Since Ψ is nondecreasing, $\Psi(\widetilde{b}) \geq \Psi(b) = k\big(\widetilde{b}, m^k(\widetilde{b})\big)$. That is,

$$k\big(\widetilde{b}, m^z(\widetilde{b})\big) \geq k\big(\widetilde{b}, m^k(\widetilde{b})\big).$$

Because $k(\widetilde{b}, \cdot)$ is strictly increasing in its second argument, it follows that $m^z(\widetilde{b}) \geq m^k(\widetilde{b})$, i.e. $\mu^z(\widetilde{b}, b, w) \geq \mu^k(\widetilde{b}, b, w)$. The case $\widetilde{b} < b$ is analogous: since Ψ is nondecreasing, we now have

$$\Psi(\widetilde{b}) \leq \Psi(b) = k\big(\widetilde{b}, m^k(\widetilde{b})\big),$$

so strict monotonicity of $k(\widetilde{b}, \cdot)$ yields $\mu^z(\widetilde{b}, b, w) \leq \mu^k(\widetilde{b}, b, w)$. *Q.E.D.*

A.3.4 Proposition 1

The first proof under the milder assumption XXX

Proof. Let (H, W) and $(\widetilde{H}, \widetilde{W})$ be independent copies of the history–talent pair. Define

$$Z \coloneqq z(H, W), \quad V \coloneqq v(H, W), \qquad \widetilde{Z} \coloneqq z(\widetilde{H}, \widetilde{W}), \quad \widetilde{V} \coloneqq v(\widetilde{H}, \widetilde{W}),$$

and assume (as in the text) that all relevant distributions are atomless so $\mathbb{P}(Z = \widetilde{Z}) = \mathbb{P}(V = \widetilde{V}) = 0$. Fundamental-mobility of career and welfare are

$$\widehat{\mathcal{M}}^z \coloneqq \mathbb{P}\big(Z > \widetilde{Z} \,\big|\, \widetilde{H} \succ H\big), \qquad \widehat{\mathcal{M}}^v \coloneqq \mathbb{P}\big(V > \widetilde{V} \,\big|\, \widetilde{H} \succ H\big),$$

Write

$$A^z \coloneqq \{Z > \widetilde{Z},\ \widetilde{H} \succ H\}, \qquad A^v \coloneqq \{V > \widetilde{V},\ \widetilde{H} \succ H\}.$$

Then

$$\widehat{\mathcal{M}}^z = \frac{\mathbb{P}(A^z)}{\mathbb{P}(\widetilde{H} \succ H)}, \qquad \widehat{\mathcal{M}}^v = \frac{\mathbb{P}(A^v)}{\mathbb{P}(\widetilde{H} \succ H)}.$$

(i) Single crossing of v over $z \Rightarrow \widehat{\mathcal{M}}^z \geq \widehat{\mathcal{M}}^v$.

Assume the single-crossing condition of v over z: for all h, h' with $h' \succ h$ and all w, w',

$$h' \succ h \quad \Rightarrow \quad \Big(v(h, w) \geq v(h', w')\Big) \implies \Big(z(h, w) > z(h', w')\Big). \tag{A5}$$

By atomlessness, $\{V > \widetilde{V}\}$ and $\{V \geq \widetilde{V}\}$ coincide almost surely. Hence, on the event $\{\widetilde{H} \succ H\}$, (A5) implies $V > \widetilde{V} \quad \Longrightarrow \quad Z > \widetilde{Z}$ a.s.

Equivalently, $A^v \subseteq A^z$ (up to a null set), so $\mathbb{P}(A^v) \leq \mathbb{P}(A^z)$ and thus

$$\widehat{\mathcal{M}}^v = \frac{\mathbb{P}(A^v)}{\mathbb{P}(\widetilde{H} \succ H)} \ \leq \ \frac{\mathbb{P}(A^z)}{\mathbb{P}(\widetilde{H} \succ H)} = \widehat{\mathcal{M}}^z.$$

If the inclusion is strict on a set of positive probability, then $\widehat{\mathcal{M}}^z > \widehat{\mathcal{M}}^v$.

(ii) Conversely, if $\widehat{\mathcal{M}}^z \geq \widehat{\mathcal{M}}^v$ for every atomless DGP, then the single-crossing condition of v over z must hold.

We prove the contrapositive. Suppose (A5) fails. Then there exist h, h' with $h' \succ h$ and talents w, w' such that

$$v(h, w) \geq v(h', w'), \qquad z(h, w) \leq z(h', w'). \tag{A6}$$

By continuity of v and z in w, we may (if necessary) perturb (w, w') slightly and assume strict inequalities

$$v(h, w) > v(h', w'), \qquad z(h, w) < z(h', w'). \tag{A7}$$

By continuity again, there exists $\varepsilon > 0$ such that for all $w' \in (w - \varepsilon, w + \varepsilon)$ and all $w'' \in (w' - \varepsilon, w' + \varepsilon)$,

$$v(h, w') > v(h', w''), \qquad z(h, w') < z(h', w''). \tag{A8}$$

Now define an atomless DGP for (H, W) as follows. Let

$$\mathbb{P}(H = h) = \mathbb{P}(H = h') = \tfrac{1}{2},$$

and conditional on H let

$$W \mid (H = h) \sim \text{Uniform}(w - \varepsilon, w + \varepsilon), \qquad W \mid (H = h') \sim \text{Uniform}(w' - \varepsilon, w' + \varepsilon),$$

truncated to $\mathbb{R}_+$ if needed. The resulting marginal distribution of W is absolutely continuous (hence atomless), and (H, W) satisfies the standing assumptions. Let (H, W) and $(\widetilde{H}, \widetilde{W})$ be i.i.d. under this DGP.

By construction,

$$\{\widetilde{H} \succ H\} = \{H = h, \widetilde{H} = h'\}, \quad \mathbb{P}(\widetilde{H} \succ H) = \tfrac{1}{4}.$$

On $\{\widetilde{H} \succ H\}$ we have $H = h, W \in (w - \varepsilon, w + \varepsilon)$ and $\widetilde{H} = h', \widetilde{W} \in (w' - \varepsilon, w' + \varepsilon)$, so (A8) implies

$$V = v(H, W) > v(\widetilde{H}, \widetilde{W}) = \widetilde{V}, \qquad Z = z(H, W) < z(\widetilde{H}, \widetilde{W}) = \widetilde{Z} \quad \text{a.s. on } \{\widetilde{H} \succ H\}.$$

Therefore,

$$\widehat{\mathcal{M}}^v = \mathbb{P}(V > \widetilde{V} \mid \widetilde{H} \succ H) = 1, \qquad \widehat{\mathcal{M}}^z = \mathbb{P}(Z > \widetilde{Z} \mid \widetilde{H} \succ H) = 0$$

for this (atomless) DGP, so $\widehat{\mathcal{M}}^z < \widehat{\mathcal{M}}^v$. This contradicts the hypothesis that $\widehat{\mathcal{M}}^z \geq \widehat{\mathcal{M}}^v$ for every history–talent DGP satisfying the model's atomlessness assumptions. Hence the single-crossing condition of v over z must hold.

Combining (i) and (ii) yields the proposition. *Q.E.D.*

The second proof of Proposition 1 UNDER STRONG ASSUMPTION and using calculus

Proof. Another proof under Assumption **??**. Step 1 We first prove (**??**) when (h,w) is arbitrary close to $\left(\widetilde{h},\widetilde{w}\right)$, i.e., (b,w) is arbitrary close to $\left(\widetilde{b},\widetilde{w}\right)$. Taylor expansion implies that (**??**) is equivalent to

$$\nabla\widehat{v}\cdot\varepsilon>0\Rightarrow\nabla\widehat{c}\cdot\varepsilon>0$$

where $\varepsilon_2>0>\varepsilon_1$. Given that $MRS^v>MRS^c$,

$$\nabla\widehat{v}\cdot\varepsilon>0\Leftrightarrow MRS^v\varepsilon_1+\varepsilon_2>0\Rightarrow MRS^c\varepsilon_1+\varepsilon_2>0$$

implying $\nabla\widehat{c}\cdot\varepsilon>0$. Step 2: Assume that $h\prec\widetilde{h}\ \&\ v(h,w)>v\left(\widetilde{h},\widetilde{w}\right)$. There exists (?) a smooth path $\gamma:[0,1]\to\mathbb{R}^{n+1}$ from point $\left(\widetilde{b},\widetilde{w}\right)$ to point (b,w), so that $b_t(t)<0$ and $\nabla\widehat{v}(\gamma(t))\cdot\gamma_t(t)>0$. Then Fundamental Theorem of Calculus implies that

$$c(h,w)-c\left(\widetilde{h},\widetilde{w}\right)=\int_0^1\nabla c(\gamma(t))\cdot\gamma_t(t)\ dt$$

Now using step 1 of the proof, we the right hand side is larger than zero. $Q.E.D.$

A.3.5 Corollary 2

Time series $\widetilde{Z}$ is an independent copy of time series Z. Let $\mathcal{Z}_t$ denote the event of $Z_t>\widetilde{Z}_t$. Given continuous marginals and independent copy, there are no ties, $\mathbb{P}\left(Z=\widetilde{Z}\right)=0$.

Define $\widetilde{H}$ and$\mathcal{H}$ for H and $\widetilde{V}$ and$\mathcal{V}$ for V in the same way. Define three concordance likelihoods:

$$M(Z):=\mathbb{P}(\mathcal{Z}\mid\neg\mathcal{Z}_{-1}),\qquad\widehat{M}(Z):=\mathbb{P}(\mathcal{Z}\mid\neg\mathcal{H}),\quad\sigma(Z):=\mathbb{P}\left(\mathcal{Z}_{-1}\,\middle|\,\neg\mathcal{H}\right).$$

We drop (Z) for now. Assume $M,\widehat{M},\sigma$ are all $<.5$, which is equivalent of positive Kendall tau betwwen Z and Z_{-1}, between Z and H and between H and Z_{-1}, respectively.

For any two events in this setting, that is

$$\{A,B\}\subseteq\{\mathcal{Z},\mathcal{Z}_{-1},\mathcal{H},\neg\mathcal{Z},\neg\mathcal{Z}_{-1},\neg\mathcal{H}\}$$

, we have $\mathbb{P}(A)=.5$, $\mathbb{P}(A|B)=\mathbb{P}(B|A)$ and $\mathbb{P}(A|B)=1-\mathbb{P}(A|\neg B)$.

$$\mathbb{P}(\mathcal{Z}\mid\neg\mathcal{Z}_{-1})=\mathbb{P}(\mathcal{Z}\mid\neg\mathcal{Z}_{-1},\mathcal{H})\,\mathbb{P}(\mathcal{H}\mid\neg\mathcal{Z}_{-1})+\mathbb{P}(\mathcal{Z}\mid\neg\mathcal{Z}_{-1},\neg\mathcal{H})\ \mathbb{P}(\neg\mathcal{H}\mid\neg\mathcal{Z}_{-1})$$

$$M=\sigma\mathbb{P}(\mathcal{Z}\mid\neg\mathcal{Z}_{-1},\mathcal{H})+(1-\sigma)\,\mathbb{P}(\mathcal{Z}\mid\neg\mathcal{Z}_{-1},\neg\mathcal{H})$$

- If $H=\beta(Z_{-1})$ a.s. with a strictly increasing β,[27] then $\sigma=0$ and $\mu=\widehat{\mu}$.

[27]This is equivalent to H and Z_{-1} have Kendall Tau of 1 or complete concordance between them (a.s. rank preserved), or events $\mathcal{Z}_{-1}=\mathcal{H}$ a.s.. If we drop the continuity/no-ties assumption, then we only get

- If there is no talent, H is sufficient statistics for Z, so

$$\mathbb{P}\left(\mathcal{Z} \mid \neg\mathcal{Z}_{-1}, \neg\mathcal{H}\right) = \mathbb{P}\left(\mathcal{Z} \mid \neg\mathcal{H}\right)$$

Then mobility is an increasing function of $\widehat{M}$ and σ as:

$$M = \sigma\left(1 - \widehat{M}\right) + (1-\sigma)\widehat{M} = \sigma + (1-2\sigma)\widehat{M} \tag{A9}$$

In the absence of talent, H is also sufficient statistics for V_{-1} on V. So we can use A9 for welfare as well.

If we assume that history is more align with welfare than career, $\sigma(Z) > \sigma(V)$, then

$$\widehat{M}(Z) > \widehat{M}(V) \Rightarrow M(Z) > M(V)$$

- Time-invariant $H = \theta$, using stationarity

$$\widehat{M}(Z) = \mathbb{P}\left(\mathcal{Z}_{-1} \mid \neg\mathcal{H}_{-1}\right) = \mathbb{P}\left(\mathcal{Z}_{-1} \mid \neg\mathcal{H}\right) = \sigma$$

$$\mu = \widehat{M}(Z)\,\mathbb{P}\left(\mathcal{Z} \mid \neg\mathcal{Z}_{-1}, \mathcal{H}\right) + \left(1 - \widehat{M}(Z)\right)\mathbb{P}\left(\mathcal{Z} \mid \neg\mathcal{Z}_{-1}, \neg\mathcal{H}\right)$$

A.3.6 Other claims

Back to the main text 1

Z1.

We want to show taht the Milgrom-Shannon single crossing k over z (Assumption 2) implies that the single crossing of v over z, that is,

$$v(h,w) \geq v\left(\widetilde{h},\widetilde{w}\right) \Rightarrow c(h,w) > c\left(\widetilde{h},\widetilde{w}\right) \tag{A10}$$

Proof by contradiction, assume $c\left(\widetilde{h},\widetilde{w}\right) \geq c(h,w)$ then the contrapositive of Assumption 2, which is

$$c\left(\widetilde{h},\widetilde{w}\right) \geq c(h,w) \Rightarrow k\left(\widetilde{h},\widetilde{w}\right) > k(h,w)$$

, implies $v\left(\widetilde{h},\widetilde{w}\right) > v(h,w)$, a contradiction.

Z2.

Proving

$$MRS^c < MRS^v \Leftrightarrow MRS^c < MRS^k$$

To see the intuition mathematically, use $v(\sigma) = U\left(c(\sigma), k(\sigma)\right)$, to rewrote the MRS of welfare

"both are increasing functions of the same latent variable", not necessarily a clean function of one variable to the other.

as:

$$\frac{\widehat{v}_b}{\widehat{v}_w} = \frac{U_c c_b + U_k k_b}{U_c c_w + U_k k_w}$$

Using this equation, we have

$$\begin{aligned} MRS^c > MRS^v &\Leftrightarrow \frac{c_b}{c_w} > \frac{\widehat{v}_b}{\widehat{v}_w} \\ &\Leftrightarrow \frac{c_b}{c_w} > \frac{U_c c_b + U_k k_b}{U_c c_w + U_k k_w} \\ &\Leftrightarrow \frac{c_b}{c_w} > \frac{k_b}{k_w} \\ &\Leftrightarrow MRS^c > MRS^k \end{aligned}$$

using the assumption that utility is non-negative with respect to b and w.

From Theory to Data

Deriving the equations used in Section 2.1.

Approach 1: Basic

Start from $\kappa(h,z) = k\big(h, \sigma^z(h,z)\big)$. Differentiating first with respect to z,

$$\partial_z \kappa(h,z) = \partial_w k(h,w)\, \sigma^z_z(h,z) = \frac{\partial_w k(h,w)}{\partial_w z(h,w)},$$

evaluated at $w = \sigma^z(h,z)$.

Now differentiate with respect to h. Since

$$\sigma^z_h(h,z) = -\frac{\partial_h z(h,w)}{\partial_w z(h,w)},$$

we obtain

$$\begin{aligned} \partial_{zh}\kappa(h,z) &= \partial_h\left(\frac{\partial_w k(h,w)}{\partial_w z(h,w)}\right) + \sigma^z_h(h,z)\,\partial_w\left(\frac{\partial_w k(h,w)}{\partial_w z(h,w)}\right) \\ &= \frac{1}{\partial_w z(h,w)}\left(\partial_{wh}k(h,w) - \partial_{ww}k(h,w)\frac{\partial_h z(h,w)}{\partial_w z(h,w)} - \partial_w k(h,w)\,\partial_w\left(\frac{\partial_h z(h,w)}{\partial_w z(h,w)}\right)\right) \\ &= \frac{1}{\partial_w z(h,w)}\Big(\partial_{ww}k(h,w)\,\Delta(h,w) + \partial_w k(h,w)\,\Delta_w(h,w)\Big), \end{aligned}$$

again evaluated at $w = \sigma^z(h,z)$.

If $\Delta(h,w) > 0$, this can be written as

$$\partial_{zh}\kappa(h,z) = \frac{\partial_w k(h,w)}{\partial_w z(h,w)}\,\Delta(h,w)\left(\frac{\partial_{ww}k(h,w)}{\partial_w k(h,w)} + \frac{\Delta_w(h,w)}{\Delta(h,w)}\right),$$

evaluated at $w = \sigma^z(h, z)$.

$$\partial_h \kappa(h, z) = \partial_h k(h, w) + \partial_w k(h, w)\, \sigma_h^z(h, z) = \partial_h k(h, w) - \partial_w k(h, w) \frac{z_h(h, w)}{z_w(h, w)} = \partial_w k(h, w)\, \Delta(h, w),$$

Using $\partial_z = \sigma_z^z\, \partial_w = \frac{1}{z_w} \partial_w$, one can obtain 10 again.

So in summary,

$$\partial_z \kappa(h, z) = \frac{\partial_w k}{\partial_w z},$$

and

$$\partial_{zh} \kappa(h, z) = \frac{1}{\partial_w z} \Big(\partial_{ww} k\, \Delta + \partial_w k\, \Delta_w \Big),$$

with both right-hand sides evaluated at $(h, w) = (h, \sigma^z(h, z))$.

Approach 2: Less Messy

For any smooth $F(h, w)$, define its pullback along σ^z by

$$(\mathcal{T} F)(h, z) := F\big(h, \sigma^z(h, z)\big).$$

Differentiating the identity $z(h, \sigma^z(h, z)) = z$ yields

$$\partial_h \sigma^z(h, z) = -\frac{\partial_h z}{\partial_w z}\big(h, \sigma^z(h, z)\big), \qquad \partial_z \sigma^z(h, z) = \frac{1}{\partial_w z\big(h, \sigma^z(h, z)\big)}.$$

Hence, for every smooth F,

$$\partial_h(\mathcal{T} F) = \mathcal{T}\left(\partial_h F - \frac{\partial_h z}{\partial_w z}\, \partial_w F \right), \qquad \partial_z(\mathcal{T} F) = \mathcal{T}\left(\frac{\partial_w F}{\partial_w z} \right).$$

Now $\kappa = \mathcal{T} k$, we obtain

$$\partial_h \kappa = \mathcal{T}\left(\partial_h k - \frac{\partial_h z}{\partial_w z}\, \partial_w k \right) = \mathcal{T}(\partial_w k\, \Delta)\,.$$

Applying ∂_z once more gives

$$\partial_{zh} \kappa = \mathcal{T}\left(\frac{\partial_w(\partial_w k\, \Delta)}{\partial_w z} \right).$$

Expanding the w-derivative,

$$\partial_w(\partial_w k\, \Delta) = \partial_{ww} k\, \Delta + \partial_w k\, \partial_w \Delta = \partial_w k \left(\partial_w \Delta + \Delta\, \frac{\partial_{ww} k}{\partial_w k} \right).$$

Therefore

$$\partial_{zh} \kappa(h, z) = \left[\frac{\partial_w k}{\partial_w z} \left(\partial_w \Delta + \Delta\, \frac{\partial_{ww} k}{\partial_w k} \right) \right]_{(h, \sigma^z(h, z))}.$$

A10

Lag-1 autocorrelation of $\hat{Z}_t$

First compute the key covariance:

$$\mathrm{Cov}(\hat{W}_t, \hat{Z}_{t-1}) = \sum_{j=0}^{\infty} \vartheta^j \,\mathrm{Cov}(\hat{W}_t, \hat{W}_{t-1-j}) = \sum_{j=0}^{\infty} \vartheta^j \rho^{1+j} = \frac{\rho}{1-\vartheta\rho}.$$

Let denote the auto-correlations as $\gamma_z(s) := \mathrm{Cov}(\hat{Z}_t, \hat{Z}_{t-s})$. From $\hat{Z}_t = \hat{W}_t + \vartheta \hat{Z}_{t-1} + \ln \alpha$,

$$\begin{aligned}
\gamma_z(0) &= \mathrm{Var}(\hat{W}_t + \vartheta \hat{Z}_{t-1}) = 1 + \vartheta^2 \gamma_z(0) + 2\vartheta \,\mathrm{Cov}(\hat{W}_t, \hat{Z}_{t-1}), \\
\gamma_z(1) &= \mathrm{Cov}(\hat{W}_t, \hat{Z}_{t-1}) + \vartheta \gamma_z(0).
\end{aligned}$$

Solving gives

$$\gamma_z(0) = \frac{1+\vartheta\rho}{(1-\vartheta^2)(1-\vartheta\rho)}, \qquad \gamma_z(1) = \frac{\vartheta+\rho}{(1-\vartheta^2)(1-\vartheta\rho)}.$$

Hence the lag-1 autocorrelation of $\hat{Z}_t$ is

$$\mathrm{Corr}(\hat{Z}_t, \hat{Z}_{t-1}) = \frac{\gamma_z(1)}{\gamma_z(0)} = \frac{\vartheta+\rho}{1+\vartheta\rho}$$

Lag-1 autocorrelation of V_t

Write $V_t = \text{const} + \alpha \hat{W}_t + \kappa \hat{Z}_{t-1}$. Then

$$\gamma_V(0) := \mathrm{Var}(V_t) = \alpha^2 + \kappa^2 \gamma_z(0) + 2\alpha\kappa \,\mathrm{Cov}(\hat{W}_t, \hat{Z}_{t-1}) = \alpha^2 + \kappa^2 \gamma_z(0) + 2\alpha\kappa \frac{\rho}{1-\vartheta\rho}.$$

where $\kappa := \alpha\vartheta + \bar{\alpha}\beta$. Also,

$$\begin{aligned}
\gamma_V(1) := \mathrm{Cov}(V_t, V_{t-1}) = \alpha^2 \mathrm{Cov}(\hat{W}_t, \hat{W}_{t-1}) + \kappa^2 \mathrm{Cov}(\hat{Z}_{t-1}, \hat{Z}_{t-2}) \\
+ \alpha\kappa \Big(\mathrm{Cov}(\hat{W}_t, \hat{Z}_{t-2}) + \mathrm{Cov}(\hat{Z}_{t-1}, \hat{W}_{t-1}) \Big).
\end{aligned}$$

Using $(\mathrm{Cov}\hat{Z}_{t-1}, \hat{Z}_{t-2}) = \gamma_z(1)$, $\mathrm{Cov}(\hat{W}_t, \hat{Z}_{t-2}) = \frac{\rho^2}{1-\vartheta\rho}$, and

$$\mathrm{Cov}(\hat{Z}_{t-1}, \hat{W}_{t-1}) = \mathrm{Cov}(\hat{Z}_t, \hat{W}_t) = \frac{1}{1-\vartheta\rho},$$

we obtain

$$\gamma_V(1) = \alpha^2\rho + \kappa^2 \gamma_z(1) + \alpha\kappa \frac{1+\rho^2}{1-\vartheta\rho}.$$

Therefore the lag-1 autocorrelation of V_t is

$$\mathrm{Corr}(V_t, V_{t-1}) = \frac{\gamma_V(1)}{\gamma_V(0)} = \frac{\alpha^2\rho + \kappa^2\gamma_z(1) + \alpha\kappa \frac{1+\rho^2}{1-\vartheta\rho}}{\alpha^2 + \kappa^2\gamma_z(0) + 2\alpha\kappa \frac{\rho}{1-\vartheta\rho}}.$$

Comparing

$$\begin{aligned}\mathrm{Corr}(V_t, V_{t-1}) - \mathrm{Corr}(\hat{Z}_t, \hat{Z}_{t-1}) &= \frac{\alpha^2\rho + \kappa^2\gamma_z(1) + \alpha\kappa\frac{1+\rho^2}{1-\vartheta\rho}}{\alpha^2 + \kappa^2\gamma_z(0) + 2\alpha\kappa\frac{\rho}{1-\vartheta\rho}} - \frac{\gamma_z(1)}{\gamma_z(0)} \\ &= \frac{\alpha^2\rho + \alpha\kappa\frac{1+\rho^2}{1-\vartheta\rho} - [\alpha^2 + 2\alpha\kappa\frac{\rho}{1-\vartheta\rho}]\frac{\gamma_z(1)}{\gamma_z(0)}}{\alpha^2 + \kappa^2\gamma_z(0) + 2\alpha\kappa\frac{\rho}{1-\vartheta\rho}}\end{aligned}$$

which can be rewritten as

$$\mathrm{Corr}(V_t, V_{t-1}) - \mathrm{Corr}(\hat{Z}_t, \hat{Z}_{t-1}) = \frac{\alpha(1-\alpha)(1-\rho^2)\,\beta}{\gamma_V(0)\,(1+\vartheta\rho)},$$

A.4 Theorem 1

Define a new one-dimensional function

$$\mu_{h,\widetilde{h}}(\cdot) := \mu(h;\widetilde{h},\cdot) \tag{A11}$$

which is $\in \mathcal{X}$ then inequality 8 can be more rigorously be written as

$$\limsup_{\delta\downarrow 0}\ \mathbb{P}\Big(W < \mu^{v}_{h,\widetilde{h}}(\widetilde{W}) \;\Big|\; |W_{-1} - \mu^{v}_{h,\widetilde{h}}(\widetilde{W}_{-1})| < \delta,\ h < \widetilde{h}\Big) > \tfrac{1}{2}.$$

Extended sign assumption

Assumption A4. *There exists a measurable function $\kappa : (0,\infty) \to (0,1]$ such that, for every map μ lying pointwise between $\mu^{v}_{h,\widetilde{h}}$ and $\mu^{z}_{h,\widetilde{h}}$ (in particular for $\mu = \mu^{r,\varepsilon}_{h,\widetilde{h}}$), we have for f-a.e. $w > 0$,*

$$\mathbb{P}(W < \mu(\widetilde{W}) \mid \widetilde{W}_{-1} = w,\ W_{-1} = \mu(w), h, \widetilde{h}) \ \geq\ \tfrac{1}{2} + \kappa(w) \tag{A12}$$

when $\tilde{h} > h$, and

$$\mathbb{P}(W < \mu(\widetilde{W}) \mid \widetilde{W}_{-1} = w,\ W_{-1} = \mu(w), h, \tilde{h}) \ \leq\ \tfrac{1}{2} - \kappa(w)$$

when $h > \tilde{h}$.

If $\mu = \mu^{v}_{h,\widetilde{h}}$ is the v-equating map from $\widetilde{h}$ to h, then $W \leq \mu^{v}_{h,\widetilde{h}}(\widetilde{W})$ is equivalent to $V \leq \widetilde{V}$, so $\Pi_{\mu^{v};h,\widetilde{h}}(t) = \mathbb{P}(\widetilde{V} > V \mid \widetilde{W}_{-1} = t,\ W_{-1} = \mu^{v}_{h,\widetilde{h}}(t), h, \widetilde{h})$. Thus Assumption A4 is a uniform extension of the strip-to-curve sign condition from μ^{v} to all interpolants between μ^{v} and μ^{z}.

With this new notation, assumption 2 is equivalent to: $\mu^{z}_{\widetilde{h},h}(w) > \mu^{v}_{\widetilde{h},h}(w)$ when $\widetilde{h} > h$ and vice versa.

A.4.1 Two Lemmas

Let

$$\mathcal{X} := \Big\{\, g : \mathbb{R}_+ \to \mathbb{R}_+ \;\Big|\; g \in C^2(0,\infty),\ g \text{ strictly increasing} \Big\}.$$

For $g \in \mathcal{X}$, define

$$\mathcal{J}(g) := \mathbb{P}\Big(W > g(\widetilde{W}),\ W_{-1} < g(\widetilde{W}_{-1})\Big), \tag{A13}$$

where (W_{-1}, W) and $(\widetilde{W}_{-1}, \widetilde{W})$ are independent copies of two consecutive draws from the talent process.

Directional derivative. Fix $\bar{\mu} \in \mathcal{X}$. For $\eta \in C^2(\mathbb{R}_+)$ with $\eta(0) = 0$, $\mathbb{E}\,|\eta(W)| < \infty$ and a tail-weight integrability condition:

$$\int_0^\infty \big(1 - F(a)\big)\,|\eta'(a)|\,da < \infty$$

Assume $\bar{\mu} + \varepsilon\eta \in \mathcal{X}$ for small $|\varepsilon|$. The Gâteaux derivative is

$$D\mathcal{J}(\bar{\mu})[\eta] := \lim_{\varepsilon \to 0} \frac{\mathcal{J}(\bar{\mu} + \varepsilon\eta) - \mathcal{J}(\bar{\mu})}{\varepsilon},$$

whenever the limit exists. $\mathcal{X}$ is constrained (not a vector space), so this is really a derivative along feasible perturbations. This is fine because we explicitly assume $\bar{\mu} + \varepsilon\eta \in \mathcal{X}$ for small $|\varepsilon|$.

Lemma A1. *The directional (Gâteaux) derivative exists and admits the representation*

$$D\mathcal{J}(\bar{\mu})[\eta] \;=\; \int_0^\infty G_{\bar{\mu}}(a)\, \eta'(a)\, da, \tag{A14}$$

where $G_{\bar{\mu}} : \mathbb{R}_+ \to \mathbb{R}$ is defined by

$$G_{\bar{\mu}}(a) := \mathbb{E}\Big[f(\bar{\mu}(W_{-1}))\big(1 - \partial_1 C(\bar{U}_{-1}, \bar{U})\big)\, \mathbf{1}_{\{W_{-1} \geq a\}} - f(\bar{\mu}(W))\, \partial_2 C(\bar{U}_{-1}, \bar{U})\, \mathbf{1}_{\{W \geq a\}} \Big], \tag{A15}$$

with $\bar{U}_t := F(\bar{\mu}(W_t))$. Moreover, $G_{\bar{\mu}}$ is absolutely continuous and, for f-a.e. a,

$$G'_{\bar{\mu}}(a) \;=\; f(a)\, f(\bar{\mu}(a))\, \big(P_1(a) + P_2(a) - 1\big), \tag{A16}$$

where

$$\begin{aligned} P_1(a) &:= \mathbb{E}[\partial_1 C(F(\bar{\mu}(a)), F(\bar{\mu}(W))) \mid W_{-1} = a], \\ P_2(a) &:= \mathbb{E}[\partial_2 C(F(\bar{\mu}(W_{-1})), F(\bar{\mu}(a))) \mid W = a]. \end{aligned}$$

The notation $\mathbb{E}[\cdot \mid W_{-1} = a]$ and $\mathbb{E}[\cdot \mid W = a]$ conditions on null events; this is fine because on $\mathbb{R}_+$ we can fix regular conditional expectation kernels, and all such identities are understood as holding for almost every a.

Proof. **Step 1: Gâteaux derivative.** Fix $x, y \in \mathbb{R}_+$. Since W_{-1} is atomless, we may freely replace strict by weak inequalities. By Sklar's theorem,

$$\begin{aligned} \mathbb{P}(W > \bar{\mu}(y),\, W_{-1} < \bar{\mu}(x)) &= \mathbb{P}(W_{-1} < \bar{\mu}(x)) - \mathbb{P}(W_{-1} \leq \bar{\mu}(x),\, W \leq \bar{\mu}(y)) \\ &= F(\bar{\mu}(x)) - C\big(F(\bar{\mu}(x)),\, F(\bar{\mu}(y))\big). \end{aligned}$$

Taking expectations with $(x, y) = (\widetilde{W}_{-1}, \widetilde{W})$ and using $(\widetilde{W}_{-1}, \widetilde{W}) \overset{d}{=} (W_{-1}, W)$ yields

$$\mathcal{J}(\bar{\mu}) = \mathbb{E}\Big[F(\bar{\mu}(W_{-1})) - C\big(F(\bar{\mu}(W_{-1})), F(\bar{\mu}(W))\big) \Big]. \tag{A17}$$

Differentiate (A17) along $\bar{\mu}_\varepsilon = \bar{\mu} + \varepsilon\eta$. We differentiate under the expectation; this is fine because f is bounded and $\partial_i C \in [0, 1]$, so the difference quotients are dominated by a constant multiple of $|\eta(W_{-1})| + |\eta(W)|$, which is integrable by assumption. By the chain rule and the standing

interchange assumptions,

$$D\mathcal{J}(\bar{\mu})[\eta] = \mathbb{E}\Big[f(\bar{\mu}(W_{-1}))\big(1-\partial_1 C(\bar{U}_{-1},\bar{U})\big)\,\eta(W_{-1}) - f(\bar{\mu}(W))\,\partial_2 C(\bar{U}_{-1},\bar{U})\,\eta(W)\Big], \tag{A18}$$

where $\bar{U}_t := F(\bar{\mu}(W_t))$.

Since $\eta(0) = 0$, for each $x \geq 0$,

$$\eta(x) = \int_0^x \eta'(a)\,da \;=\; \int_0^\infty \mathbf{1}_{\{x\geq a\}}\,\eta'(a)\,da.$$

Substitute this representation for $\eta(W_{-1})$ and $\eta(W)$ in (A18) and apply Fubini/Tonelli: Tonelli needs nonnegativity but η' can change sign; this is fine because the tail-weight condition $\int_0^\infty (1-F(a))|\eta'(a)|da < \infty$ gives absolute integrability, so plain Fubini applies.

$$D\mathcal{J}(\bar{\mu})[\eta] = \int_0^\infty \mathbb{E}\Big[f(\bar{\mu}(W_{-1}))\big(1-\partial_1 C(\bar{U}_{-1},\bar{U})\big)\,\mathbf{1}_{\{W_{-1}\geq a\}} - f(\bar{\mu}(W))\,\partial_2 C(\bar{U}_{-1},\bar{U})\,\mathbf{1}_{\{W\geq a\}}\Big]\,\eta'(a)\,da,$$

which is (A14) with $G_{\bar{\mu}}$ defined by (A15).

Step 2: Derivative of $G_{\bar{\mu}}$. Let, for $(w_{-1}, w) \in \mathbb{R}_+^2$,

$$\begin{aligned} A(w_{-1},w) &:= f(\bar{\mu}(w_{-1}))\Big(1-\partial_1 C\big(F(\bar{\mu}(w_{-1})),F(\bar{\mu}(w))\big)\Big),\\ B(w_{-1},w) &:= f(\bar{\mu}(w))\,\partial_2 C\big(F(\bar{\mu}(w_{-1})),F(\bar{\mu}(w))\big). \end{aligned}$$

Then

$$G_{\bar{\mu}}(a) = \mathbb{E}\big[A(W_{-1},W)\mathbf{1}_{\{W_{-1}\geq a\}}\big] - \mathbb{E}\big[B(W_{-1},W)\mathbf{1}_{\{W\geq a\}}\big]\,.$$

For any integrable $\psi(X,Y)$ with X atomless and density f_X, the map $a \mapsto \mathbb{E}[\psi(X,Y)\mathbf{1}_{\{X\geq a\}}]$ is absolutely continuous with

$$\frac{d}{da}\mathbb{E}[\psi(X,Y)\mathbf{1}_{\{X\geq a\}}] = -f_X(a)\,\mathbb{E}[\psi(a,Y)\mid X=a] \quad \text{for } f_X\text{-a.e. } a.$$

To see that, note that with a density we can write the expectation as $\int_a^\infty m(x)f_X(x)\,dx$ where $m(x) = \mathbb{E}[\psi(x,Y)\mid X=x]$, and then apply the fundamental theorem of calculus.

Applying this with $X = W_{-1}$ and $X = W$ (stationarity gives $f_{W_{-1}} = f_W = f$), we obtain for f-a.e. a,

$$\begin{aligned} G'_{\bar{\mu}}(a) &= -f(a)\,\mathbb{E}[A(a,W)\mid W_{-1}=a] + f(a)\,\mathbb{E}[B(W_{-1},a)\mid W=a]\\ &= f(a)f(\bar{\mu}(a))\Big[\mathbb{E}[\partial_1 C(F(\bar{\mu}(a)),F(\bar{\mu}(W)))\mid W_{-1}=a]\\ &\quad + \mathbb{E}[\partial_2 C(F(\bar{\mu}(W_{-1})),F(\bar{\mu}(a)))\mid W=a] - 1\Big]. \end{aligned} \tag{A19}$$

We state the derivative identity "for f-a.e. a" rather than Lebesgue-a.e.; this is fine because the

right-hand side always carries a factor $f(a)$, so any differences on $\{f = 0\}$ never matter in uses of the formula. $Q.E.D.$

Lemma A2. *Fix $a \geq 0$ and let $\bar{\mu} \in \mathcal{X}$. Then:*

1. *The function $t \mapsto G_{\bar{\mu}}(t)$ is absolutely continuous on $[a, \infty)$; hence it is differentiable for almost every $t \geq a$ in the Lebesgue sense (in particular, for f-almost every t).*
2. $\lim_{t\to\infty} G_{\bar{\mu}}(t) = 0$.
3. $\int_a^\infty |G'_{\bar{\mu}}(t)|\, dt < \infty$.
4. *If $\bar{\mu} = \bar{\mu}_{H,\widetilde{H}}$ is a random element of $\mathcal{X}$, measurable with respect to $\sigma(H, \widetilde{H})$ and satisfying $\bar{\mu}_{H,\widetilde{H}} \in \mathcal{X}$ a.s., then*
$$\int_a^\infty \mathbb{E}_{H,\widetilde{H}}\big[|G'_{\bar{\mu}_{H,\widetilde{H}}}(t)|\big]\, dt \;\leq\; \|f\|_\infty\,(1 - F(a)) < \infty.$$

Proof. **(1)** Recall from Lemma A1 (Step 2) that

$$G_{\bar{\mu}}(t) = \mathbb{E}\big[A(W_{-1}, W)\mathbf{1}_{\{W_{-1}\geq t\}}\big] - \mathbb{E}\big[B(W_{-1}, W)\mathbf{1}_{\{W\geq t\}}\big],$$

where A, B are the (nonnegative) functions defined there. Using $0 \leq \partial_1 C, \partial_2 C \leq 1$ on $(0,1)^2$ and $\|f\|_\infty < \infty$, we have, for all $(w_{-1}, w) \in \mathbb{R}^2_+$,

$$0 \leq A(w_{-1}, w) \leq f(\bar{\mu}(w_{-1})) \leq \|f\|_\infty, \qquad 0 \leq B(w_{-1}, w) \leq f(\bar{\mu}(w)) \leq \|f\|_\infty.$$

Hence $A(W_{-1}, W)$ and $B(W_{-1}, W)$ are integrable. By disintegration with respect to the marginal densities of W_{-1} and W,

$$\mathbb{E}\big[A(W_{-1}, W)\mathbf{1}_{\{W_{-1}\geq t\}}\big] = \int_t^\infty \mathbb{E}[A(s, W) \mid W_{-1} = s]\; f(s)\, ds,$$
$$\mathbb{E}\big[B(W_{-1}, W)\mathbf{1}_{\{W\geq t\}}\big] = \int_t^\infty \mathbb{E}[B(W_{-1}, s) \mid W = s]\; f(s)\, ds.$$

Define

$$\alpha_{\bar{\mu}}(s) := \mathbb{E}[A(s, W) \mid W_{-1} = s]\; f(s), \qquad \beta_{\bar{\mu}}(s) := \mathbb{E}[B(W_{-1}, s) \mid W = s]\; f(s).$$

Then, for $t \geq a$,

$$G_{\bar{\mu}}(t) = \int_t^\infty \big(\alpha_{\bar{\mu}}(s) - \beta_{\bar{\mu}}(s)\big)\, ds.$$

Moreover, $0 \leq \alpha_{\bar{\mu}}(s) \leq \|f\|_\infty f(s)$ and $0 \leq \beta_{\bar{\mu}}(s) \leq \|f\|_\infty f(s)$, so $\alpha_{\bar{\mu}}, \beta_{\bar{\mu}} \in L^1([a, \infty))$ and therefore $\alpha_{\bar{\mu}} - \beta_{\bar{\mu}} \in L^1([a, \infty))$. It follows that $G_{\bar{\mu}}$ is absolutely continuous on $[a, \infty)$ with

$$G'_{\bar{\mu}}(t) = -(\alpha_{\bar{\mu}}(t) - \beta_{\bar{\mu}}(t)) \quad \text{for Lebesgue-a.e. } t \geq a.$$

(2) Since $\alpha_{\bar{\mu}} - \beta_{\bar{\mu}} \in L^1([a,\infty))$,

$$\lim_{t\to\infty} G_{\bar{\mu}}(t) = \lim_{t\to\infty} \int_t^\infty (\alpha_{\bar{\mu}}(s) - \beta_{\bar{\mu}}(s))\, ds = 0,$$

e.g. by dominated convergence applied to $(\alpha_{\bar{\mu}} - \beta_{\bar{\mu}})\mathbf{1}_{\{s\geq t\}}$.

(3) For Lebesgue-a.e. $t \geq a$,

$$|G'_{\bar{\mu}}(t)| = |\alpha_{\bar{\mu}}(t) - \beta_{\bar{\mu}}(t)| \leq \max\{\alpha_{\bar{\mu}}(t), \beta_{\bar{\mu}}(t)\} \leq \|f\|_\infty f(t).$$

Integrating gives

$$\int_a^\infty |G'_{\bar{\mu}}(t)|\, dt \leq \|f\|_\infty \int_a^\infty f(t)\, dt = \|f\|_\infty (1 - F(a)) < \infty.$$

(4) The pointwise bound in (3) holds for each realization of $\bar{\mu}_{H,\widetilde{H}}$, hence for Lebesgue-a.e. $t \geq a$,

$$\mathbb{E}_{H,\widetilde{H}}\big[|G'_{\bar{\mu}_{H,\widetilde{H}}}(t)|\big] \leq \|f\|_\infty f(t).$$

Integrating over $t \in [a,\infty)$ and applying Tonelli's theorem yields

$$\int_a^\infty \mathbb{E}_{H,\widetilde{H}}\big[|G'_{\bar{\mu}_{H,\widetilde{H}}}(t)|\big]\, dt \leq \|f\|_\infty \int_a^\infty f(t)\, dt = \|f\|_\infty (1 - F(a)) < \infty.$$

$Q.E.D.$

A.4.2 The Proof of Theorem 1

Proof. For a random map $\bar{\mu}_{H,\widetilde{H}} \in \mathcal{X}$ and given the definition of $\mathcal{J}(g)$ from A13, define

$$\bar{\mathcal{J}}(\bar{\mu}_{H,\widetilde{H}}) := \mathbb{E}_{H,\widetilde{H}}\left[\mathcal{J}\big(\bar{\mu}_{H,\widetilde{H}}\big)\right], \tag{A20}$$

By monotonicity of $y(h,\cdot)$ and the definition of the equating map $\mu^y(h;\tilde{h},\cdot)$, the event $\{Y > \widetilde{Y},\ Y_{-1} < \widetilde{Y}_{-1}\}$ is equivalent to $\{W > \mu^y_{H,\widetilde{H}}(\widetilde{W}),\ W_{-1} < \mu^y_{H,\widetilde{H}}(\widetilde{W}_{-1})\}$. Since $\mathbb{P}(Y_{-1} < \widetilde{Y}_{-1}) = 1/2$ (i.i.d. and atomless), we have

$$\mathcal{M}^y = 2\,\bar{\mathcal{J}}(\mu^y_{H,\widetilde{H}}). \tag{A21}$$

Thus $\mathcal{M}^z > \mathcal{M}^v$ is equivalent to

$$\bar{\mathcal{J}}(\mu^z_{H,\widetilde{H}}) > \bar{\mathcal{J}}(\mu^v_{H,\widetilde{H}}).$$

We will prove this in four steps.

Step 1: Defining and Deriving the ironing path. Define the nonnegative gap

$$\mu^{\Delta}_{h,\widetilde{h}}(w) := \left|\mu^{z}_{h,\widetilde{h}}(w) - \mu^{v}_{h,\widetilde{h}}(w)\right| \geq 0.$$

Assumption 5 implies $\mu^{\Delta}_{h,\widetilde{h}}$ is strictly increasing on $\mathbb{R}_+$, hence $\partial_w \mu^{\Delta}_{h,\widetilde{h}}(w) \geq 0$ for all $w > 0$ and $\partial_w \mu^{\Delta}_{h,\widetilde{h}}(w) > 0$ for Lebesgue-a.e. $w > 0$.

Under Assumption 2, for each history pair $(h, \widetilde{h})$ with $h \neq \widetilde{h}$, we can define

$$s(h, \tilde{h}) := \operatorname{sign}(\mu^{z}_{h,\widetilde{h}}(w) - \mu^{v}_{h,\widetilde{h}}(w)) = \operatorname{sign}(h - \tilde{h}) \in \{+1, -1\}.$$

such that for any $w \geq 0$, we have

$$\mu^{z}_{h,\widetilde{h}}(w) = \mu^{v}_{h,\widetilde{h}}(w) + s(h, \widetilde{h})\, \mu^{\Delta}_{h,\widetilde{h}}(w). \tag{A22}$$

Fix $\varepsilon > 0$. Let $\chi \in C^{\infty}(\mathbb{R})$ be a smooth cutoff function such that $\chi(a) = 1$ for $a \leq -1$, $\chi(a) = 0$ for $a \geq 0$, and χ is nonincreasing, so that it transitions from 1 to 0 on $[-1, 0]$. Define $\rho(a) := -\chi'(a) \geq 0$. Then $\int_{\mathbb{R}} \rho(a)\, da = 1$. Moreover, $\chi\big(\frac{a-r}{\varepsilon}\big)$ is nonincreasing in a, equals 1 for $a \leq r - \varepsilon$, and equals 0 for $a \geq r$; hence it transitions from 1 to 0 as a increases from $r - \varepsilon$ to r.

For $r \geq 0$ and $w \geq 0$, recall the (history-dependent) ironing family

$$\mu^{r,\varepsilon}_{h,\widetilde{h}}(w) := \mu^{v}_{h,\widetilde{h}}(w) + s(h, \widetilde{h}) \int_0^w \chi\Big(\frac{a - r}{\varepsilon}\Big)\, \partial_w \mu^{\Delta}_{h,\widetilde{h}}(a)\, da. \tag{A23}$$

In one limit case, we have $\lim_{r \to 0^+} \mu^{r,\varepsilon}_{h,\widetilde{h}}(w) = \mu^{v}_{h,\widetilde{h}}(w)$, and in the other

$$\lim_{r \to \infty} \mu^{r,\varepsilon}_{h,\widetilde{h}}(w) = \mu^{z}_{h,\widetilde{h}}(w) - s(h, \widetilde{h}) \mu^{\Delta}_{h,\widetilde{h}}(0).$$

If $\mu^{\Delta}_{h,\widetilde{h}}(0) = 0$, then $\lim_{r \to \infty} \mu^{r,\varepsilon}_{h,\widetilde{h}}(w) = \mu^{z}_{h,\widetilde{h}}(w)$. [did we assume that?]

Since $\mu^{\Delta} \in C^2(0, \infty)$ and $\chi \in C^{\infty}$, the definition of $\mu^{r,\varepsilon}_{h,\widetilde{h}}$ immediately gives $\mu^{r,\varepsilon}_{h,\widetilde{h}} \in C^2(0, \infty)$. If $s(h, \widetilde{h}) = +1$, then $\mu^{r,\varepsilon}(w) \geq \mu^{v}(w) \geq 0$, and for $0 \leq x < y$,

$$\mu^{r,\varepsilon}(y) - \mu^{r,\varepsilon}(x) = \mu^{v}(y) - \mu^{v}(x) + \int_x^y \chi((a - r)/\varepsilon)(\mu^{\Delta})'(a)\, da > 0.$$

The case of $s(h, \widetilde{h}) = -1$ leads similarly to $\mu^{r,\varepsilon}(w) \geq \mu^{z}(w) \geq 0$ and $\mu^{r,\varepsilon}(y) - \mu^{r,\varepsilon}(x) > 0$. Hence $\mu^{r,\varepsilon}_{h,\widetilde{h}}$ is C^2 and strictly increasing, so $\mu^{r,\varepsilon}_{h,\widetilde{h}} \in \mathcal{X}$; if $h = \widetilde{h}$, then trivially $\mu^{r,\varepsilon}_{h,h} = \mu^{v}_{h,h} \in \mathcal{X}$.

Then, for all $w > 0$,

$$\begin{aligned}
\partial_r \mu^{r,\varepsilon}_{h,\widetilde{h}}(w) &= s(h, \widetilde{h}) \int_0^w \partial_r \left[\chi\Big(\frac{a - r}{\varepsilon}\Big)\right] \partial_w \mu^{\Delta}_{h,\widetilde{h}}(a)\, da \\
&= \frac{s(h, \widetilde{h})}{\varepsilon} \int_0^w \rho\Big(\frac{a - r}{\varepsilon}\Big)\, \partial_w \mu^{\Delta}_{h,\widetilde{h}}(a)\, da,
\end{aligned} \tag{A24}$$

and hence

$$\partial_w \partial_r \mu^{r,\varepsilon}_{h,\widetilde{h}}(w) = \frac{s(h,\widetilde{h})}{\varepsilon}\rho\Big(\frac{w-r}{\varepsilon}\Big)\,\partial_w \mu^{\Delta}_{h,\widetilde{h}}(w). \tag{A25}$$

Step 2: Differentiating $\bar{\mathcal{J}}(\mu^{r,\varepsilon}_{H,\widetilde{H}})$.
Using the definition A20, we have

$$\bar{\mathcal{J}}(\mu^{r,\varepsilon}_{H,\widetilde{H}}) = \mathbb{E}_{H,\widetilde{H}}\Big[\mathcal{J}\big(\mu^{r,\varepsilon}_{H,\widetilde{H}}\big)\Big].$$

Since, by construction, for $P_{H,\widetilde{H}}$-a.e. $(h,\widetilde{h})$ the map $r \mapsto \mu^{r,\varepsilon}_{h,\widetilde{h}}(w)$ is pathwise C^∞ for every $w \geq 0$, it is enough to assume that Lemma A1 applies pathwise at $\mu^{r,\varepsilon}_{H,\widetilde{H}}$ in the direction $\partial_r \mu^{r,\varepsilon}_{H,\widetilde{H}}$, and that the resulting pathwise derivative admits a local L^1-dominating envelope, so that differentiation may be passed through the outer expectation. Hence $r \mapsto \bar{\mathcal{J}}(\mu^{r,\varepsilon}_{H,\widetilde{H}})$ is locally absolutely continuous and for Lebesgue-a.e. $r > 0$,

$$\partial_r \bar{\mathcal{J}}(\bar{\mu}^{r,\varepsilon}_{H,\widetilde{H}}) = \mathbb{E}_{H,\widetilde{H}}\Big[D\mathcal{J}(\mu^{r,\varepsilon}_{H,\widetilde{H}})\big[\partial_r \mu^{r,\varepsilon}_{H,\widetilde{H}}\big]\Big] \tag{A26}$$

Now using Lemma A1, we have:

$$\partial_r \bar{\mathcal{J}}(\bar{\mu}^{r,\varepsilon}_{H,\widetilde{H}}) = \mathbb{E}_{H,\widetilde{H}}\left[\int_0^\infty G_{\mu^{r,\varepsilon}_{H,\widetilde{H}}}(a)\,\partial^2_{r,w}\mu^{r,\varepsilon}_{H,\widetilde{H}}(a)\,da\right]. \tag{A27}$$

Now using A25,

$$\partial_r \bar{\mathcal{J}}(\bar{\mu}^{r,\varepsilon}_{H,\widetilde{H}}) = \mathbb{E}_{H,\widetilde{H}}\left[\int_{r-\varepsilon}^{r} G_{\mu^{r,\varepsilon}_{H,\widetilde{H}}}(a)\,\frac{s(H,\widetilde{H})}{\varepsilon}\rho\Big(\frac{a-r}{\varepsilon}\Big)\,\partial_w \mu^{\Delta}_{H,\widetilde{H}}(a)\,da\right]. \tag{A28}$$

The terms $\rho\Big(\frac{a-r}{\varepsilon}\Big)$ and $\partial_w \mu^{\Delta}_{H,\widetilde{H}}(a)$ are positive. The next step will investigate the sign of $G_{\mu^{r,\varepsilon}_{H,\widetilde{H}}}(a)\frac{s(H,\widetilde{H})}{\varepsilon}$.

Step 3: Sign of $G_{\mu^{r,\varepsilon}_{H,\widetilde{H}}}(a)\,s(H,\widetilde{H})$.

Fix a history pair $(H,\widetilde{H})$ and fix a deterministic map $\mu \in \mathcal{X}$ (later we will take $\mu = \mu^{r,\varepsilon}_{H,\widetilde{H}}$). Let (W_{-1},W) and $(\widetilde{W}_{-1},\widetilde{W})$ be independent copies of two consecutive talents.

For $t > 0$, define

$$\Pi_{\mu;H,\widetilde{H}}(t) := \mathbb{P}\Big(W \leq \mu(\widetilde{W}) \,\Big|\, \widetilde{W}_{-1} = t,\ W_{-1} = \mu(t),\ H,\widetilde{H}\Big), \tag{A29}$$

where conditioning on equalities is understood via fixed regular conditional distributions (equiv-

alently, as a strip-to-curve limit).

Using the conditional CDF representation $\mathbb{P}(W \le y \mid W_{-1} = x) = \partial_1 C(F(x), F(y))$ and independence of the two copies, we have for f-a.e. $t > 0$,

$$\begin{aligned}\Pi_{\mu;H,\widetilde{H}}(t) &= \mathbb{E}\Big[\mathbb{P}\Big(W \le \mu(\widetilde{W}) \mid W_{-1} = \mu(t), \widetilde{W}\Big) \,\Big|\, \widetilde{W}_{-1} = t,\ H, \widetilde{H}\Big] \\ &= \mathbb{E}\Big[\partial_1 C\Big(F(\mu(t)), F(\mu(\widetilde{W}))\Big) \,\Big|\, \widetilde{W}_{-1} = t,\ H, \widetilde{H}\Big] \\ &= \mathbb{E}[\partial_1 C(F(\mu(t)), F(\mu(W))) \,|\, W_{-1} = t]\,. \end{aligned} \tag{A30}$$

Now to get the derivative of G_μ, assume lag-1 pairwise reversibility of talent and copula symmetry, i.e. $(W_{-1}, W) \overset{d}{=} (W, W_{-1})$ and $C(u, v) = C(v, u)$. Then Lemma A1 (Step 2) implies that for f-a.e. $t > 0$,

$$G'_{\mu;H,\widetilde{H}}(t) = f(t)\, f(\mu(t))\Big(2\,\Pi_{\mu;H,\widetilde{H}}(t) - 1\Big). \tag{A31}$$

By Lemma A2(ii), $\lim_{x\to\infty} G_{\mu;H,\widetilde{H}}(x) = 0$, hence for all $a \ge 0$,

$$G_{\mu;H,\widetilde{H}}(a) = -\int_a^\infty f(t)\, f(\mu(t))\Big(2\,\Pi_{\mu;H,\widetilde{H}}(t) - 1\Big)\, dt. \tag{A32}$$

Rewrite Assumption A4 using the $s(.)$ and $\Pi(.)$ functions:

$$-s(h, \widetilde{h})\left[\Pi_{\mu;h,\widetilde{h}}(w) - \tfrac{1}{2}\right] \ge\ \kappa(w) \tag{A33}$$

Combining (A32) with (A33) yields, for all $a \ge 0$,

$$\begin{aligned} s(H, \widetilde{H})\, G_{\mu;H,\widetilde{H}}(a) &= \int_a^\infty f(t)\, f(\mu(t))\Big[-s(H, \widetilde{H})\big(2\,\Pi_{\mu;H,\widetilde{H}}(t) - 1\big)\Big]\, dt \\ &\ge 2\int_a^\infty f(t)\, f(\mu(t))\, \kappa(t)\, dt\ >\ 0. \end{aligned} \tag{A34}$$

In particular, taking $\mu = \mu^{r,\varepsilon}_{H,\widetilde{H}}$ gives for all $a \ge 0$,

$$s(H, \widetilde{H})\, G_{\mu^{r,\varepsilon}_{H,\widetilde{H}}}(a)\ >\ 0. \tag{A35}$$

Step 4: Monotonicity of $\bar{\mathcal{J}}(\bar{\mu}^{r,\varepsilon}_{H,\widetilde{H}})$ with respect to r .

Using (A28), $\rho \ge 0$, and the sign conclusion from Step 3, namely (A35), we obtain for Lebesgue-a.e. $r > 0$,

$$\partial_r \bar{\mathcal{J}}(\bar{\mu}^{r,\varepsilon}_{H,\widetilde{H}}) = \mathbb{E}_{H,\widetilde{H}}\left[\int_{\max\{0, r-\varepsilon\}}^{r} \Big(s(H, \widetilde{H})\, G_{\mu^{r,\varepsilon}_{H,\widetilde{H}}}(a)\Big)\, \frac{1}{\varepsilon}\rho\Big(\frac{a-r}{\varepsilon}\Big)\, \partial_w \mu^{\Delta}_{H,\widetilde{H}}(a)\, da\right]\ \ge\ 0.$$

Moreover, under Assumption 5 together with the differentiability used in the construction (e.g. $\Delta^{\sharp}_{H,\widetilde{H}} \in C^1$ and strictly increasing, hence $\int_{r-\varepsilon}^{r} \Delta^{\sharp\prime}_{H,\widetilde{H}}(a)\, da = \Delta^{\sharp}_{H,\widetilde{H}}(r) - \Delta^{\sharp}_{H,\widetilde{H}}(r-\varepsilon) > 0$), and

choosing χ strictly decreasing on $(-1,0)$ so that $\rho > 0$ on $(-1,0)$, the inner integral is strictly positive whenever $H \neq \widetilde{H}$. Since $\mathbb{P}(H \neq \widetilde{H}) > 0$ in the non-degenerate case, it follows that

$$\partial_r \bar{\mathcal{J}}(\bar{\mu}^{r,\varepsilon}_{H,\widetilde{H}}) \; > \; 0 \quad \text{for Lebesgue-a.e. } r > 0.$$

Therefore $\bar{\mathcal{J}}(\bar{\mu}^{r,\varepsilon}_{H,\widetilde{H}})$ is strictly increasing in r on $(0,\infty)$, and hence

$$\bar{\mathcal{J}}(\bar{\mu}^{0,\varepsilon}_{H,\widetilde{H}}) < \lim_{r\to\infty} \bar{\mathcal{J}}(\bar{\mu}^{r,\varepsilon}_{H,\widetilde{H}}). \tag{A36}$$

By dominated convergence and pointwise convergence of $\mu^{r,\varepsilon}_{H,\widetilde{H}}$ to $\mu^{z}_{H,\widetilde{H}}$, (A36) yields

$$\mathbb{E}_{H,\widetilde{H}}\left[\mathcal{J}(\mu^{v}_{H,\widetilde{H}})\right] < \mathbb{E}_{H,\widetilde{H}}\left[\mathcal{J}(\mu^{z}_{H,\widetilde{H}})\right].$$

$Q.E.D.$

A.5 Case Studies

A.5.1 Time-Invariant Background: Binary Case

In this section, we study the case in which background H is Bernoulli, $H \in \{0,1\}$, time-invariant, and non-degenerate. We build on our results from the general time-invariant case in Section XXX to obtain results in the binary case, useful for our... in Section XXX. The next section presents the main results, while all proofs are collected in Section A.5.1.

Let

$$q \; \coloneqq \; \mathbb{P}(H=1)\,\mathbb{P}(H=0) \; > \; 0,$$

so that both background types occur with positive probability.

Results

Recall the definition of the talent metric in equation 5, where $\mu^z(\widetilde{h}; h, w)$ is the level of w under background $\widetilde{h}$ that reproduces the same career outcome as (h,w). In the binary-h case, the only nontrivial mappings are between $h=0$ and $h=1$. So we only need $\mu^z(0;1,w) = \mu^z_{0,1}(w)$, see equation A11, which we denote by $\mu^z(w)$ for simplification. so that, by construction,

$$z(1,w) = z(0,\mu^z(w)).$$

With this new notation, the assumption 2, i.e. $MRS^v > MRS^z$ is equivalent to

$$\mu^v(w) > \mu^z(w).$$

It becomes useful to define the talent metric for talent itself: $\mu^w(x) = x$

For both $g \in \{z, v\}$, define new r.v. $T_t^g \equiv F\mu^g(W_t)$. The difference between them is denoted by $\Delta_t \equiv T_t^v - T_t^z$. Moreover, $T_t^w = F_W(W_t) =: U_t$ is the probability integral transform (PIT) uniformly distributed, denoted by U_t.

The next result formalizes the intuition that more talent is required to compensate for disadvantaged history, i.e. higher the mean of T, the lower is semi-mobility.

Lemma 1. *The semi-mobility can be written as* $\widehat{\mathcal{M}}^g = 1 - \mathbb{E}(T^g)$ *for* $g \in \{z, v\}$.

A simple implication of this lemma is that by symmetry of $(W, \widetilde{W})$, we have:

$$\mathbb{P}\left(Z > \widetilde{Z} \mid H > \widetilde{H}\right) = \mathbb{P}\left(\mu^z(W) > \widetilde{W}\right) = \mathbb{E}[T^z] \tag{A37}$$

Claim A2. *For a fixed talent distribution* W*, the semi-mobility of career is larger than welfare if and only if* $\mathbb{E}(\Delta_t) > 0$,

$$\widehat{\mathcal{M}}^z > \widehat{\mathcal{M}}^v \Leftrightarrow \mathbb{E}(\Delta_t) > 0$$

One sufficient condition to get $\mathbb{E}(\Delta_t) > 0$ is to assume that $\mu^v(W)$ first-order stochastically dominates $\mu^z(W)$, strict domination on a set of positive measure gives the strict inequality. The first-order stochastically dominant implies that for any increasing ϕ, $\mathbb{E}[\phi(\mu^v(W))] \geq \mathbb{E}[\phi(\mu^z(W))]$. With $\phi = F_W$, $\mathbb{E}(T^v) \geq \mathbb{E}(T^z)$. Now, we will study ...

Corollary 1. *Assumption 2 is necessary and sufficient condition for* $\widehat{\mathcal{M}}^z > \widehat{\mathcal{M}}^v$ *for any distribution* W.

Under assumption 2, we have $\mu^v(w) > \mu^z(w) > w$,which implies $T_t^v > T_t^z > U_t$ a.s. hence $\mathbb{E}[T_t^v] > \mathbb{E}[T_t^z] > \mathbb{E}[F_W(W_t)] = \frac{1}{2}$, as $F_W(W_t)$ is uniform over $[0, 1]$. If background does not matter $\mu^z(x) \to x$ a.s., then $\mathbb{E}[T_t^z] \to 0.5$.

Lemma 2. $\mathcal{M}(W) = \chi(F_W)$

Lemma 3. *Mobility in income is weighted average of talent mobility and ...*

$$\mathcal{M}^z = (1 - 2q)\,\mathcal{M}(W) \;+\; 2q\,\mathbb{E}\left[T^z - C_W\left(T_{-1}^z, T^z\right)\right] \tag{A38}$$

or equivalently,

$$\mathcal{M}^z = (1 - 2q)\,\mathcal{J}(I) \;+\; 2q\,\mathcal{J}(\mu^z)$$

Proofs

Proof. of Lemma 1

Given the binary nature of background, we know that

$$\widehat{\mathcal{M}}^z = \mathbb{P}\left(z(0, W) > z\left(1, \widetilde{W}\right) \mid \widetilde{H} = 1, H = 0\right)$$

Using the definition of talent metric, i.e., $z\left(1,\widetilde{W}\right)=z\left(0,\mu^{z}\left(\widetilde{W}\right)\right)$, we can write

$$\begin{aligned}\widehat{\mathcal{M}}^{z} &= \mathbb{P}\left(Z>\widetilde{Z}\,|\,\widetilde{H}>H\right)\\ &= \mathbb{P}\left(z\left(0,W\right)>z\left(0,\mu^{z}\left(\widetilde{W}\right)\right)\right)\\ &= \mathbb{P}\left(W>\mu^{z}\left(\widetilde{W}\right)\right)\end{aligned}$$

By iterated expectations and using independence of W and $\widetilde{W}$ and equation , we have

$$\mathbb{P}\left(W>\mu^{z}\left(\widetilde{W}\right)\right)=\mathbb{E}\Big[\mathbb{P}\Big(W>\mu^{z}\left(\widetilde{W}\right)\mid\widetilde{W}\Big)\Big]=\mathbb{E}\Big[\overline{F}_{W}(\mu^{z}\left(\widetilde{W}\right))\Big]=1-\mathbb{E}\left[T_{t}^{z}\right] \tag{A39}$$

$Q.E.D.$

Proof. of Claim A2

Using Lemma 1, and its equivalence for V, we have

$$\widehat{\mathcal{M}}^{z}>\widehat{\mathcal{M}}\left(V\right)\Leftrightarrow\mathbb{E}\left(T^{z}\right)<\mathbb{E}\left(T^{v}\right)\Leftrightarrow\mathbb{E}\left(\Delta_{t}\right)>0$$

$Q.E.D.$

Proof. of Corollary 1

($\Rightarrow$) By Claim A2, $\widehat{\mathcal{M}}^{z}>\widehat{\mathcal{M}}\left(V\right)$ is equivalent to $\mathbb{E}\left(\Delta_{t}\right)>0$. Assumption 2 gives $T^{v}>T^{z}$ a.s., hence $\mathbb{E}(\Delta_{t})>0$. ($\Leftarrow$) Using Claim A2, $\mathbb{E}\left(\Delta_{t}\right)>0$ holds

$$\mathbb{E}[F_{W}(\mu^{v}(W))]\geq\mathbb{E}[F_{W}(\mu^{z}(W))]$$

for all F_W. So ... $Q.E.D.$

Proof. of Corollary **??**

As $W,\widetilde{W}$ are independent and let $g:\mathbb{R}\to\mathbb{R}$ be Borel-measurable, then

$$\mathbb{P}(W\leq g(\widetilde{W}))=\mathbb{E}\Big[\mathbb{P}(W\leq g(Y)\mid\widetilde{W})\Big]=\mathbb{E}[F_{W}(g(\widetilde{W}))], \tag{A40}$$

and similarly,

$$\mathbb{E}[g\left(W\right)]=\mathbb{E}_{W}\left[\mathbb{P}\left(\widetilde{W}\leq F^{-1}g\left(W\right)\right)\right]=\mathbb{P}\left(F\left(\widetilde{W}\right)\leq g\left(W\right)\right)$$

This is the same idea used in the proof of Lemma 1. In the same vein, as (W_{-1},W) is independent of $(\widetilde{W}_{-1},\widetilde{W})$, then

$$\mathbb{P}(W_{-1}\leq g(\widetilde{W}_{-1}),\,W\leq g(\widetilde{W}))=\mathbb{E}\Big[F_{W_{-1},W}(g(\widetilde{W}_{-1}),g(\widetilde{W}))\Big] \tag{A41}$$

and

$$\mathbb{P}(F_{W}(W_{-1})\leq g(\widetilde{W}_{-1}),\,F_{W}(W)\leq g(\widetilde{W}))=\mathbb{E}\Big[C_{W}(g(\widetilde{W}_{-1}),g(\widetilde{W}))\Big] \tag{A42}$$

So

$$\begin{aligned}\chi(g) &= \mathbb{P}\left(F_W\left(\widetilde{W}\right) \leq g(W)\right) - \mathbb{P}(F_W(W_{-1}) \leq g(\widetilde{W}_{-1}),\, F_W(W) \leq g(\widetilde{W})) \\ &= \mathbb{P}(F_W(W_{-1}) > g(\widetilde{W}_{-1}),\, F_W(W) \leq g(\widetilde{W}))\end{aligned}$$

$Q.E.D.$

Proof. of Lemma 2

Given that $F_W(W)$ is uniformly distributed and using the definition of $\chi(.)$, we have

$$\begin{aligned}\chi(F_W) &= \mathbb{E}\left[F_W(W) - C_W\left(F_W(W_{-1}), F_W(W)\right)\right] \\ &= \frac{1}{2} - \mathbb{E}\left[C_W\left(F_W(W_{-1}), F_W(W)\right)\right]\end{aligned}$$

which is exactly the definition of $\mathcal{M}(W)$, [Refer to kendal section]. $Q.E.D.$

Proof. of Lemma 3

Step 1. We will appeal to our results from XXX. To start given our definition that, for any $(h, \widetilde{h}) \in \{0,1\}^2$,

$$\Omega^z(h,\widetilde{h}) := \mathbb{P}\Big(W > \mu^z\Big(\widetilde{h}; h, \widetilde{W}\Big)\ ,\ \mu^z\Big(\widetilde{h}; h, \widetilde{W}_{-1}\Big) > W_{-1}\Big),$$

where the probability is taken over the joint law of $(W_{-1}, W, \widetilde{W}_{-1}, \widetilde{W})$, with $(\widetilde{W}_{-1}, \widetilde{W}) \overset{d}{=} (W_{-1}, W)$ and $(\widetilde{W}_{-1}, \widetilde{W}) \perp (W_{-1}, W)$.

Because $\mu^z(h; h, w) = w$, the same-background cases satisfy

$$\Omega^z(0,0) = \Omega^z(1,1) = \mathbb{P}\Big(W > \widetilde{W}\,,\ \widetilde{W}_{-1} > W_{-1}\Big) = \chi(F_W),$$

Using the binary-h notation $\mu^z(w) \equiv \mu^z(0; 1, w)$, we obtain the cross-background term

$$\Omega^z(1,0) = \mathbb{P}\Big(W > \mu^z\Big(\widetilde{W}\Big)\ ,\ \mu^z\Big(\widetilde{W}_{-1}\Big) > W_{-1}\Big).$$

Moreover, the time reversibility of talent time series implies that $\Omega^z(0,1) = \Omega^z(0,1)$.

$$\Omega^z(0,1) = \mathbb{P}\Big(\mu^z(W) > \widetilde{W}\,,\ \widetilde{W}_{-1} > \mu^z(W_{-1})\Big)$$

Step 2. Let $A \equiv \left\{W < \mu^z\left(\widetilde{W}\right)\right\}$ and $B \equiv \left\{W_{-1} < \mu^z\left(\widetilde{W}_{-1}\right)\right\}$

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}\left(W < \mu^z\left(\widetilde{W}\right),\, W_{-1} < \mu^z\left(\widetilde{W}_{-1}\right)\right) \\
&= \mathbb{E}F_{W_{-1},W}\left(\mu^z\left(\widetilde{W}_{t-1}\right), \mu^z\left(\widetilde{W}\right)\right) \\
&= \mathbb{E}C_W\left(F_W\left(\mu^z\left(\widetilde{W}_{t-1}\right)\right), F_W\left(\mu^z\left(\widetilde{W}_t\right)\right)\right) \\
&= \mathbb{E}C_W\left(T^z_{-1}, T^z\right) \qquad \text{(A43)}
\end{aligned}$$

where we used the independence of time series XXX and XXX (see equation A41).

By stationarity and Lemma 1 (precisely equation A37), $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{E}(T^z)$. Given the symmetry of

$$\Omega^z(1,0) + \Omega^z(0,1) = \mathbb{P}\left(W > \mu^z\left(\widetilde{W}\right), W_{-1} < \mu^z\left(\widetilde{W}_{-1}\right)\right) + \mathbb{P}\left(\mu^z(W) > \widetilde{W}, \mu^z(W_{-1}) < \widetilde{W}_{-1}\right)$$

Hence,

$$\begin{aligned}
\Omega^z(1,0) + \Omega^z(0,1) &= \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B^c) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B) \\
&= 2\left[\mathbb{E}(T^z) - \mathbb{E}C_W\left(T^z_{-1}, T^z\right)\right]
\end{aligned}$$

Step 3: Now using our result XXXX, the mobility statistic averages over background types:

$$\mathcal{M}^z = \mathbb{E}_{H,\widetilde{H}}\left[\Omega^z(H, \widetilde{H})\right].$$

then

$$\mathcal{M}^z = (1-q)\,\chi(F_W) + q\Big(\Omega^z(1,0) + \Omega^z(0,1)\Big).$$

$Q.E.D.$

A.5.2 Parental Career as background $h = z_{-1}$

This is the detils of the simulation in 1.1.

Setup. Let the state (family background) be talent and parental career $\sigma = (w, z_{-1})$, with $w > 0$ and $z_{-1} > 0$.

Preferences over (z, k) are modeled by Stone-Geary utility:

$$u(z,k) = \alpha \ln(z - \gamma_z) + \bar{\alpha} \ln(k - \gamma_k), \qquad \alpha \in (0,1),\ \bar{\alpha} \coloneqq 1 - \alpha.$$

where γ_z and γ_k are subsistence (minimum required) quantities of z and k, respectively.

The agent chooses inputs $x, y \in \mathbb{R}$ subject to $x + y \leq T$ (with $T \in \mathbb{R}$), and production technologies of professional and personal lifes are $z = F_z(x, \sigma), k = F_k(y, \sigma)$. The value function is

$$v(\sigma) = \max_{x,y} \alpha \ln(z - \gamma_z) + \bar{\alpha} \ln(k - \gamma_k) \quad \text{s.t.} \quad x + y \leq T,\ z = F_z(x, \sigma),\ k = F_k(y, \sigma).$$

Assume

$$F_z(x, \sigma) = wxz_{-1}^{\vartheta}, \quad F_k(y, \sigma) = yz_{-1}^{\beta},$$

where $\vartheta, \beta \in \mathbb{R}_+$, capturing the role of background in productions of each dimension of life. Given the strictly increasign utility and $y = T - x$, the constraints imply $z = wxz_{-1}^{\vartheta}, k = (T - x)z_{-1}^{\beta}$. Equivalently, using $x = \frac{z}{wz_{-1}^{\vartheta}}$, the induced frontier (optimal personal-life function) is

$$\kappa(z, z_{-1}, w) \equiv \left[T - \frac{z}{wz_{-1}^{\vartheta}}\right] z_{-1}^{\beta}.$$

The problem reduces to a one-dimensional choice in x:

$$\max_{x} \alpha \ln\big(wxz_{-1}^{\vartheta} - \gamma_z\big) + \bar{\alpha} \ln\big((T - x)z_{-1}^{\beta} - \gamma_k\big).$$

The interior FOC is

$$\frac{\alpha w z_{-1}^{\vartheta}}{wxz_{-1}^{\vartheta} - \gamma_z} = \frac{\bar{\alpha} z_{-1}^{\beta}}{(T - x)z_{-1}^{\beta} - \gamma_k}.$$

Solving yields

$$x^*(\sigma) = \alpha T - \alpha \gamma_k z_{-1}^{-\beta} + \bar{\alpha} \frac{\gamma_z}{wz_{-1}^{\vartheta}}, \qquad y^*(\sigma) = T - x^*(\sigma).$$

Hence the optimal outputs are

$$z(\sigma) = wx^*(\sigma)z_{-1}^{\vartheta} = \alpha \left(T - \frac{\gamma_k}{z_{-1}^{\beta}}\right) wz_{-1}^{\vartheta} + \bar{\alpha}\, \gamma_z,$$

$$k(\sigma) = y^*(\sigma)z_{-1}^{\beta} = \alpha\, \gamma_k + \bar{\alpha} \left[T - \frac{\gamma_z}{wz_{-1}^{\vartheta}}\right] z_{-1}^{\beta}.$$

The logs require $z(\sigma) > \gamma_z$ and $k(\sigma) > \gamma_k$, which is equivalent to

$$\mathcal{J}(\sigma) \coloneqq T - \frac{\gamma_k}{z_{-1}^{\beta}} - \frac{\gamma_z}{wz_{-1}^{\vartheta}} > 0.$$

One can think of this term as a "surplus" term, which serves as a sufficient statistic for the interiority of the solution.

We can rewrite the optimal outputs levels using $\mathcal{J}$ as

$$z(\sigma) - \gamma_z = \alpha\, wz_{-1}^{\vartheta} \mathcal{J}(\sigma), \qquad k(\sigma) - \gamma_k = \bar{\alpha}\, z_{-1}^{\beta} \mathcal{J}(\sigma),$$

and the indirect utility as

$$v(\sigma) = \alpha \ln\big(\alpha w z_{-1}^{\vartheta} \mathcal{J}(\sigma)\big) + \bar{\alpha} \ln\big(\bar{\alpha} z_{-1}^{\beta} \mathcal{J}(\sigma)\big),$$

or, expanded,

$$v(\sigma) = \ln \mathcal{J}(\sigma) + \alpha \ln w + \big(\alpha\vartheta + \bar{\alpha}\beta\big) \ln z_{-1} + \alpha \ln \alpha + \bar{\alpha} \ln \bar{\alpha}.$$

Comparative statics. Taking partial derivatives of the policy functions $z(\sigma)$ and $k(\sigma)$, with respect to w and z_{-1},

$$\frac{\partial z}{\partial w} = \alpha z_{-1}^{\vartheta} \left(T - \gamma_k z_{-1}^{-\beta}\right), \qquad \frac{\partial z}{\partial z_{-1}} = \alpha w \left(\vartheta T z_{-1}^{\vartheta-1} - (\vartheta - \beta)\gamma_k z_{-1}^{\vartheta-\beta-1}\right).$$

$$\frac{\partial k}{\partial w} = \bar{\alpha} \frac{\gamma_z}{w^2} z_{-1}^{\beta-\vartheta}, \qquad \frac{\partial k}{\partial z_{-1}} = \bar{\alpha} \left(\beta T z_{-1}^{\beta-1} - (\beta - \vartheta)\frac{\gamma_z}{w} z_{-1}^{\beta-\vartheta-1}\right).$$

MRSs. (Comparative statics ratios)

For $\partial z/\partial w \neq 0$ (equivalently $T - \gamma_k z_{-1}^{-\beta} \neq 0$),

$$MRS^z = -\frac{\dfrac{\partial z}{\partial z_{-1}}}{\dfrac{\partial z}{\partial w}} = \frac{w}{z_{-1}} \frac{(\vartheta - \beta)\gamma_k - \vartheta T z_{-1}^{\beta}}{T z_{-1}^{\beta} - \gamma_k}$$

and for $\partial k/\partial w \neq 0$ (equivalently $\gamma_z > 0$),

$$MRS^k = -\frac{\dfrac{\partial k}{\partial z_{-1}}}{\dfrac{\partial k}{\partial w}} = \frac{w}{z_{-1}} \left[\beta - \vartheta - \beta \frac{T w z_{-1}^{\vartheta}}{\gamma_z}\right]$$

The difference between the two MRSs can be rewritten as

$$MRS^z - MRS^k = \frac{w^2}{\gamma_z z_{-1}} \frac{\beta T z_{-1}^{\vartheta}}{T - \gamma_k z_{-1}^{-\beta}} \mathcal{J}(\sigma)$$

Therefore, career MRS is above family, $MRS^z > MRS^k$ for any interior solution (i.e. when $\mathcal{J}(\sigma) > 0$). More precisely, $MRS^z \geq MRS^k \iff \mathcal{J}(\sigma) > 0$, where equality holds when $\beta = 0$.

Using the expressions we derived,

$$\frac{\dfrac{\partial k}{\partial w}}{\dfrac{\partial z}{\partial w}} = \frac{\bar{\alpha}}{\alpha} \frac{\gamma_z}{w^2} \frac{z_{-1}^{2(\beta-\vartheta)}}{T z_{-1}^{\beta} - \gamma_k}.$$

Special case (Cobb–Douglas+Log-Normal). Assume W_t is jointly standardized log-normal and denote its lag-1 autocorrelation by ρ. Set $\gamma_z = \gamma_k = 0$ and normalize the constant T to 1. Then

$$\ln z = \ln(\alpha) + \ln w + \vartheta \ln z_{-1}, \qquad v = \alpha \ln \alpha + \bar{\alpha} \ln \bar{\alpha} + \alpha \ln w + \left(\alpha\vartheta + \bar{\alpha}\beta\right) \ln z_{-1}.$$

Iterating the recursion for $\ln z_t$ yields, for any $t \geq 0$,

$$\ln Z_t = \frac{\ln(\alpha)}{1-\vartheta} + \sum_{j=0,1,\ldots} \vartheta^j \ln W_{t-j},$$

Assuming that W_t, then also is Z_t. Given that

$$V = \bar{\alpha} \ln \bar{\alpha} + \alpha \ln Z + \bar{\alpha}\beta \ln Z_{-1}$$

, we can show that the lag-1 autocorrelation of $\ln z_t$

$$\mathrm{Corr}(\ln Z_t, \ln Z_{t-1}) = \frac{\vartheta + \rho}{1 + \vartheta\rho}$$

and

$$\mathrm{Corr}(V_t, V_{t-1}) = \frac{\alpha^2\rho(1-\vartheta\rho) + \kappa^2 \frac{\vartheta+\rho}{1-\vartheta^2} + \alpha\kappa(1+\rho^2)}{\alpha^2(1-\vartheta\rho) + \kappa^2 \frac{1+\vartheta\rho}{1-\vartheta^2} + 2\alpha\kappa\rho}.$$

where $\kappa \coloneqq \alpha\vartheta + \bar{\alpha}\beta$, and the latter is larger if $\alpha \in (0,1)$, $\rho < 1$, and $\beta > 0$. See calculation in A.3.6, where we show that

$$\mathrm{Corr}(V_t, V_{t-1}) - \mathrm{Corr}(\hat{Z}_t, \hat{Z}_{t-1}) = \frac{\alpha(1-\alpha)(1-\rho^2)\,\beta}{\gamma_V(0)\,(1+\vartheta\rho)},$$

[The intuition is as follows. Given the absence of the income effect, the intergenerational mobility in income rank is $\frac{6}{\pi}\arcsin\left(\frac{\rho}{2}\right)$, because $c = \alpha w$.[28] But the intergenerational mobility in welfare is different as the welfare rank depends on the ranking of a function of own and parents' productivity, more precisely

$$\alpha \ln(w) + \overline{\alpha}\beta \ln(\tilde{w})\,.$$

For the case where $\gamma_c, \gamma_k > 0$, we use a simulation. The results are depicted in Figure **??** below and the details are relegated to the Appendix. The parameter ρ is set so that income Spearman rank correlation matches the data. xxx]

Simulation: Stone-geary +Log-Normal n dynasties, burnin T, so we consider (z_{t-1}, z_t) and (v_{t-1}, v_t) for $t \in \{T, T+1, ..., T+M\}$

We use $n = 1$, we could have use another does not matter as T is large enough

[28] This follows from the known relationship between Pearson's and Spearman's correlation in the bivariate normal case; see (**?**, Proposition 5.1.3) for example.

In simulations it is convenient to compute

$$v(\sigma) = \alpha \ln\big(z(\sigma) - \gamma_z\big) + \bar{\alpha} \ln\left(\bar{\alpha}\left(\left[T - \frac{\gamma_z}{w z_{-1}^{\vartheta}}\right] z_{-1}^{\beta} - \gamma_k\right)\right),$$

Draw w_0 and z_0 from standard log-normal. Technical note on how to draw in the simulation: To have interior solutions, logs are well-defined, we need to have $w > 1$ and $z > 1$. The latter can be simpified to $z_{-1} > \left(\frac{w}{w-1}\right)^{1/\beta}$. So the sufficient condition is that in each step we have $w > 1 + \frac{1}{z_{-1}^{\beta}-1}$.

$\alpha = 0.5$, $\gamma_k = \gamma_z = 1$

A.5.3 Markovian joint log-normal DGP for talent

We denote the joint CDF of standard normals with correlation ρ with $\Phi_\rho(.,.)$ and the CDF of standard normal with $\Phi(.)$. So the former is the cdf of $(\ln W_{-1}, \ln W_t)$ and the latter the cdf of .... Note that talent series are time-reversible in this case.

Proof of time-invariant case, Claim 3 With a time-invariant background, we can use our results and provide an analytical ....

Assume a multiplicative shift talent metric function: $\mu(x) = ax$, where $a > 1$. Using an independent $X \sim N(0,1)$,

$$\begin{aligned} \mathbb{E}\big[F_W(\mu(W_t)\big] &= \mathbb{E}\big[\Phi(\ln(hW_t))\big] \\ &= \mathbb{P}\big(X \leq \ln(hW_t)\big) \\ &= \mathbb{P}\big(X - \ln W_t \leq \ln a\big) \end{aligned} \tag{A44}$$

Since $X - \ln W_t \sim N(0,2)$, then

$$\mathbb{E}\big[F_W(\mu(W_t)\big] = \Phi\Big(\frac{\ln a}{\sqrt{2}}\Big), \tag{A45}$$

The bivariate version of the same trick is as follows: With an independent $(X_1, X_2) \sim N\big(0,0;1,1;\rho\big)$, we have

$$\begin{aligned} \mathbb{E}\big[F_{W_{-1},W}(\mu(W_{t-1}), \mu(W_t))\big] &= \mathbb{E}\big[\Phi_\rho(\ln(hW_{t-1}), \ln(hW_t))\big] \\ &= \mathbb{P}(X_1 - \ln W_{t-1} \leq \ln a,\ X_2 - \ln W_t \leq \ln a) \\ &= \Phi_\rho\Big(\tfrac{\ln a}{\sqrt{2}}, \tfrac{\ln a}{\sqrt{2}}\Big), \end{aligned} \tag{A46}$$

because $(X_1 - \ln W_{t-1}, ; X_2 - W_t) \sim N\big(0,0;2,2;\rho\big)$.

Hence, an exact closed form for $\Omega(F_w\mu)$ is:

$$\Omega(F_w\mu) = \Phi\left(\frac{\ln a}{\sqrt{2}}\right) - \Phi_\rho\left(\frac{\ln a}{\sqrt{2}}, \frac{\ln a}{\sqrt{2}}\right), \tag{A47}$$

We can also exploit our proof of the theorem XXX to show the Ω, expression A47. See A.5.3

Ω is decreasing in a. To see that, we differentiate it:

$$\frac{\partial}{\partial a}\Omega(F_w\mu) = \frac{\phi\left(\frac{\ln a}{\sqrt{2}}\right)}{a\sqrt{2}}\left[1 - 2\,\Phi\left(\frac{\ln a}{\sqrt{2}}\sqrt{\frac{1-\rho}{1+\rho}}\right)\right].$$

Since $\rho > 0$ we have $\sqrt{\frac{1-\rho}{1+\rho}} \in (0,1)$, hence the argument of Φ is positive, and because Φ is strictly increasing with $\Phi(0) = \frac{1}{2}$, it follows that $\Phi(\cdot) > \frac{1}{2}$. Therefore, the bracket $1 - 2\Phi(\cdot)$ is negative, while the prefactor is positive. Consequently $\Omega' < 0$ for all $1 > \rho > 0$ and $a > 1$.

In addition, Ω is increasing in ρ as xxx, which is expected (as xxx), and the slope of Ω is decreasing in ρ, so that Ω is constant for $\rho = 1$. These are depicted in the following figure:

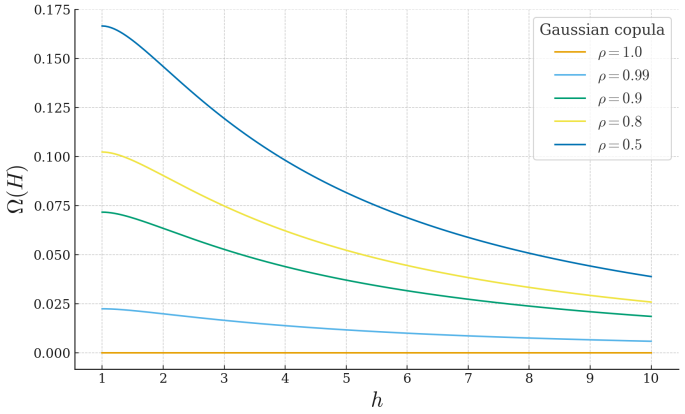


Figure A2: Theoretical

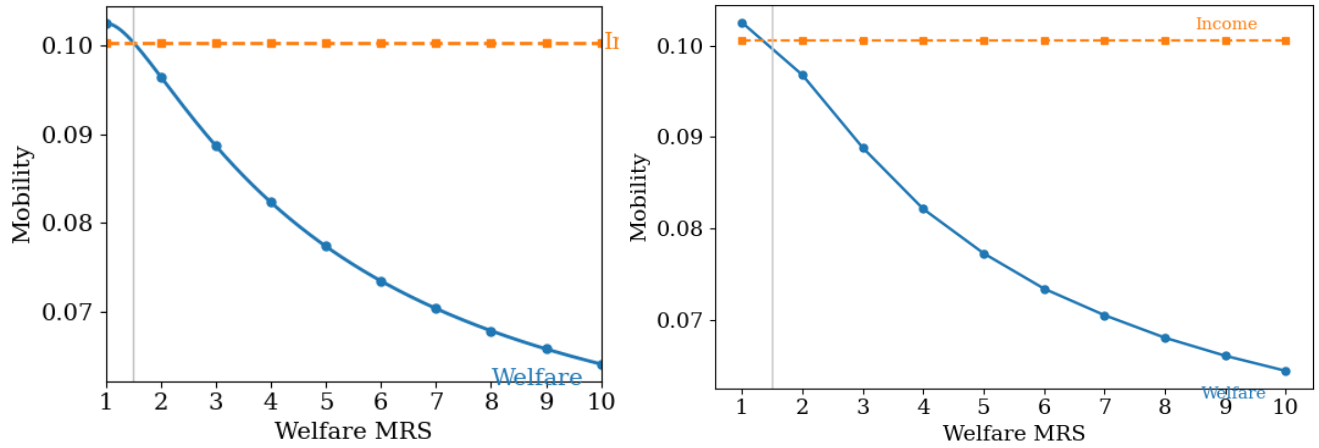
Note: This figure . more precseily, it plots $\Omega\left(F_W \circ \mu\right)$ with a multiplicative shift talent metric function: $\mu(x) = hx$.

Figure A3 Panel a compare the welfare vs income mobility for $a^c = 1.5$, $\rho = .8$ and $q = .5$

In the next step, we replicate and validate our analytical result using a simulation. Rather than imposing the analytical expression, we compute mobility directly in the simulated data using Kendall's τ. The simulation proceeds as follows. We will We can also simulate this model instead of using our expressions for Ω and M. For example $M\left(V\right) = \frac{1-\tau(V)}{4}$ where $\tau\left(V\right)$ is the Kendall tau of (V_t, V_{t-1})

Figure A3 Panel b

Figure A3: Analytical result vs. Simulation

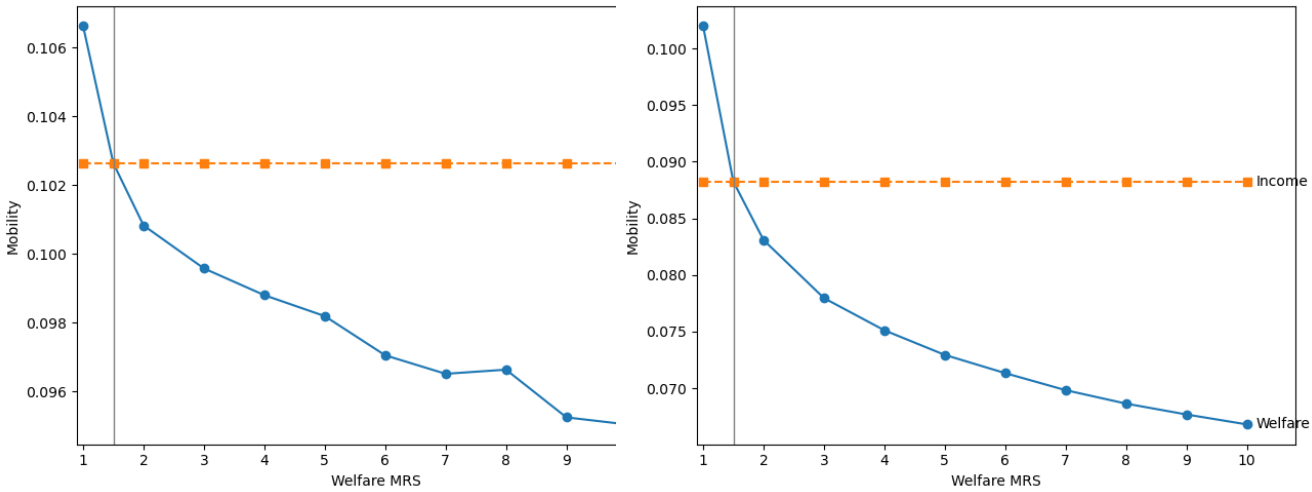


(a) Analytical (b) Simulation

Note: ...against welfare MRS (h^v.

Time-variant Background Figure A4a shows H_t is a binary indicator that is 1 if Z_{t-1} is above median. Panel b has a model where histpry has two dimensions z and v, and For Y : H_t is a binary indicator of whether Z_{t-1} is above the median. For V : H_t is a binary indicator of whether V_{t-1} is above the median.

Figure A4: Two Model Simulations



(a) Privilege by income (b) Privilege by income and welfare

Proof of claim XXX

Start with ...

$$\Psi(u) = \Phi\big(\Phi^{-1}(u) + \ln h\big), \qquad \Psi'(u) = \frac{\phi(\Phi^{-1}(u) + \ln h)}{\phi(\Phi^{-1}(u))} = \exp\Big(- \Phi^{-1}(u)\, \ln h - \tfrac{1}{2}(\ln h)^2\Big).$$

Gaussian copula partials:

$$\partial_1 C(u,v) = \Phi\left(\frac{\Phi^{-1}(v) - \rho\,\Phi^{-1}(u)}{\sigma}\right), \quad \partial_2 C(u,v) = \Phi\left(\frac{\Phi^{-1}(u) - \rho\,\Phi^{-1}(v)}{\sigma}\right).$$

Let $t := \ln h$ and $\sigma := \sqrt{1-\rho^2}$.

$$\Psi'(V) = e^{-tZ_2 - \frac{1}{2}t^2}.$$

Let $Z_1 = \Phi^{-1}(U),\ Z_2 = \Phi^{-1}(V) \overset{iid}{\sim} N(0,1)$. Then

$$\partial_2 C(\Psi(U), \Psi(V)) = \Phi\left(\frac{Z_1 - \rho Z_2 + (1-\rho)t}{\sigma}\right), \qquad \partial_1 C(U,V) = \Phi\left(\frac{Z_2 - \rho Z_1}{\sigma}\right).$$

and

$$\Psi'(V)\,\partial_2 C(\Psi(U), \Psi(V))\,\partial_1 C(U,V) = e^{-Z_2 t - \frac{1}{2}(\ln h)^2}\,\Phi\left(\frac{Z_1 - \rho Z_2 + (1-\rho)t}{\sigma}\right)\,\Phi\left(\frac{Z_2 - \rho Z_1}{\sigma}\right).$$

Let $\hat{Z}_1, \hat{Z}_2 \overset{iid}{\sim} N(0,1)$ be independent of (Z_1, Z_2). Using $\Phi(x) = \mathbb{P}(N \le x)$, we have

$$e^{-tZ_2 - \frac{1}{2}t^2}\ \mathbf{1}\Big\{\sigma\hat{Z}_1 + \rho Z_2 - Z_1 \le (1-\rho)t\Big\}\,\mathbf{1}\Big\{\sigma\hat{Z}_2 + \rho Z_1 - Z_2 \le 0\Big\}$$

Esscher tilt in Z_2. Define a new measure $\mathbb{Q}$ via

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-tZ_2 - \frac{1}{2}t^2}.$$

Then $Z_2 \sim N(-t, 1)$ under $\mathbb{Q}$, while $Z_1, \hat{Z}_1, \hat{Z}_2$ remain standard normal and independent. Consequently

$$E = \mathbb{Q}\Big(\sigma\hat{Z}_1 + \rho Z_2 - Z_1 \le (1-\rho)t,\ \sigma\hat{Z}_2 + \rho Z_1 - Z_2 \le 0\Big).$$

Let $X := \sigma\hat{Z}_1 + \rho Z_2 - Z_1$ and $Y := \sigma\hat{Z}_2 + \rho Z_1 - Z_2$. Under $\mathbb{Q}$: $\mathbb{E}_{\mathbb{Q}}[X] = -\rho t, \mathbb{E}_{\mathbb{Q}}[Y] = t, \mathrm{Var}(X) = \mathrm{Var}(Y) = 2, \mathrm{Cov}(X,Y) = -2\rho$. Adjusting for the center and scale: $X' := \frac{X+\rho t}{\sqrt{2}}$ and $Y' := \frac{Y-t}{\sqrt{2}}$. Then (X', Y') is bivariate standard normal with correlation $\mathrm{Corr}(X', Y') = -\rho$. The two inequalities transform as

$$X \le (1-\rho)t \iff X' \le \frac{t}{\sqrt{2}}, \qquad Y \le 0 \iff Y' \le -\frac{t}{\sqrt{2}}.$$

Therefore

$$\mathbb{Q}\left(X' \le \frac{t}{\sqrt{2}},\ Y' \le -\frac{t}{\sqrt{2}}\right) = \Phi_2\left(\frac{t}{\sqrt{2}}, -\frac{t}{\sqrt{2}}; -\rho\right).$$

Finally, recall $t = \ln h$. Hence

$$\mathbb{E}\big[\Psi'(V)\,\partial_2 C(\Psi(U),\Psi(V))\,\partial_1 C(U,V)\big] = \Phi_2\left(\frac{\ln h}{\sqrt{2}}, -\frac{\ln h}{\sqrt{2}}; -\rho\right).$$

B Models and Codes

Code For First Two Figures (theoretical, no simualtions) Figure A2 Figure A3 Panel a

Model 1

H Bernoulli, with stationary $P(H=1) = .5$

$$Z_t = W_t 1_{H=0} + a^z W_t 1_{H=1}$$
$$V_t = W_t 1_{H=0} + a^v W_t 1_{H=1}$$

(W_t, W_{t-1}) is jointly lognormal with parameters ($\mu_1 = \mu_2 = 0$), ($\sigma_1 = \sigma_2 = 1$), and log-correlation ($\rho = 0.8$).

$$M(Y) = \frac{1-\tau(Y)}{4}$$
$$M(V) = \frac{1-\tau(V)}{4}$$

where $\tau(Y)$ is the Kendall tau of (Z_t, Z_{t-1}) and $\tau(V)$ is the Kendall tau of (V_t, V_{t-1})

Assume $a^z = 1.5$. Plot $M(V)$ and $M(Y)$ against a^v where a^v varies from 1 to 10.

B.1 Model 2

We have a stationary setting. (W_t, W_{t-1}) is jointly lognormal with parameters ($\mu_1 = \mu_2 = 0$), ($\sigma_1 = \sigma_2 = 1$), and log-correlation ($\rho = 0.8$).

$$Z_t = W_t 1_{H_t=0} + a^z W_t 1_{H_t=1}$$
$$V_t = W_t 1_{H_t=0} + a^v W_t 1_{H_t=1}$$

and H_t is a binary indicator that is 1 if Z_{t-1} is above median of $F_Y(Z_{t-1}) > .5$ where F_Y is the CDF of V.

$$M(Y) = \frac{1-\tau(Y)}{4}$$
$$M(V) = \frac{1-\tau(V)}{4}$$

where $\tau(Y)$ is the Kendall tau of (Z_t, Z_{t-1}) and $\tau(V)$ is the Kendall tau of (V_t, V_{t-1})

Assume $a^z = 1.5$. Plot $M(V)$ and $M(Y)$ against a^v where a^v varies from 1,2,...,10 in addition to 1.5.

In the Figure, label the x-axis 1,2,...,10. X-axis title: "Welfare MRS". Y-axis title: "Mobility". Label in the figure M\left(V\right) "Welfare" and M\left(Y\right) "Income". Don't use legend. Add a vertical line at x=1.5

B.2 Model 3

We have a stationary setting. (W_t, W_{t-1}) is jointly lognormal with parameters ($\mu_1 = \mu_2 = 0$), ($\sigma_1 = \sigma_2 = 1$), and log-correlation (ρ).

$$Z_t = W_t 1_{F_Y(Z_{t-1})<s} + a^z W_t 1_{F_Y(Z_{t-1})>s}$$
$$V_t = W_t 1_{F_V(V_{t-1})<s} + a^v W_t 1_{F_V(V_{t-1})>s}$$

where the binary indicators are 1 if for example V_{t-1} is above median of V distribution (F_V is the stationary CDF of V).

and

$$M(Y) = \frac{1-\tau(Y)}{4}$$
$$M(V) = \frac{1-\tau(V)}{4}$$

where $\tau(Y)$ is the Kendall tau of (Z_t, Z_{t-1}) and $\tau(V)$ is the Kendall tau of (V_t, V_{t-1})

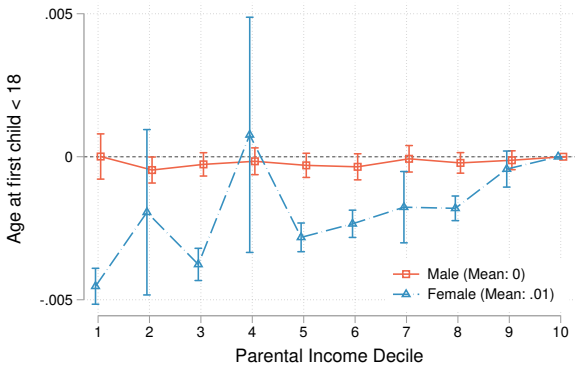
Plot $M(V)$ and $M(Y)$ against a^v where a^v varies from 1,2,...,10 in addition to 1.5.

In the Figure, label the x-axis 1,2,...,10. X-axis title: "Welfare MRS". Y-axis title: "Mobility". Label in the figure M\left(V\right) "Welfare" and M\left(Y\right) "Income". Don't use legend. Add a vertical line at x=1.5

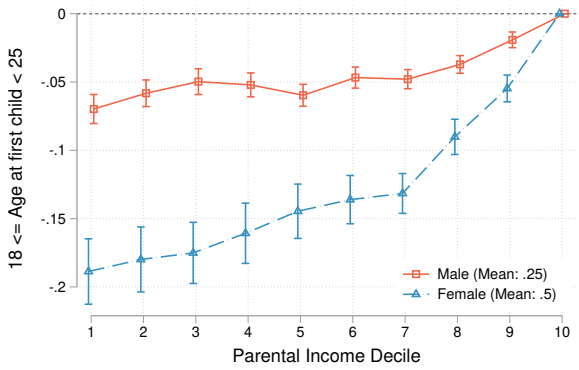
Assume $a^z = 1.5$, $\rho = .8$ and $s = .5$.

C Appendix Figures

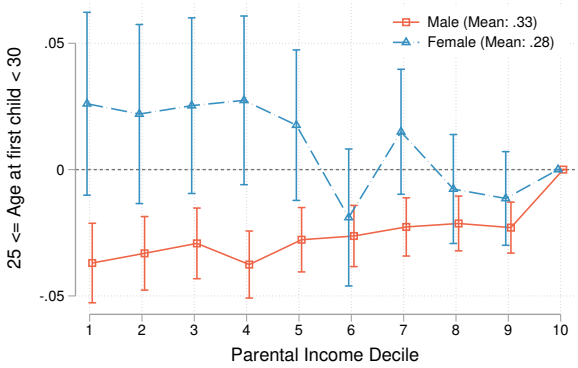
Figure C1: Age at First Birth, Non-parametric Analysis



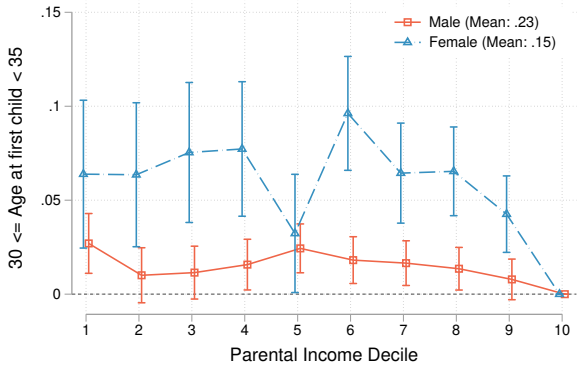
(a) Age at First Birth below 18



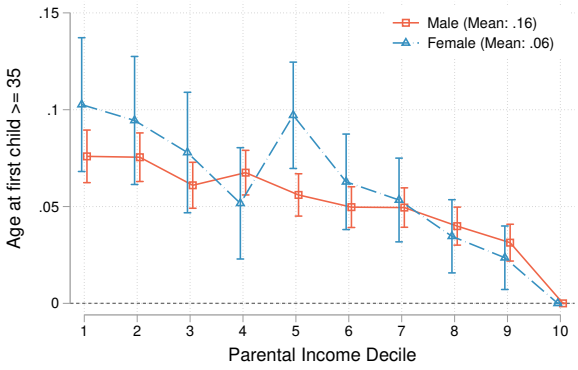
(b) Age at First Birth between 18 and 24



(c) Age at First Birth between 25 and 29



(d) Age at First Birth between 30 and 34



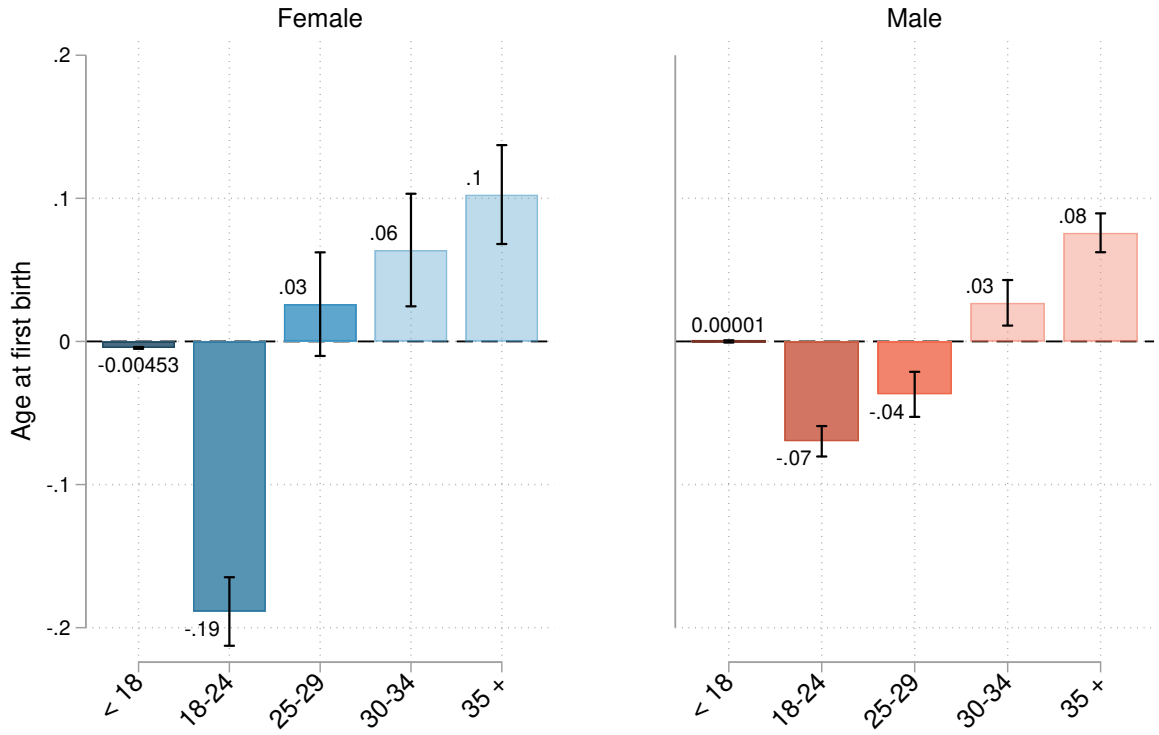
(e) Age at First Birth above 35

Note: This figure illustrates the difference in the age at the first birth between individuals at the top vs. the bottom of lifetime income by parental income deciles, using top-decile parents as the reference group. It plots the coefficients of the interaction terms, $\{\beta_i\}$, in the regression of family outcome k on a gender-specific non-parametric functions of their own lifetime income c and their parental lifetime income $\tilde{c}$:

$$k = f(c) + g(\tilde{c}) + \sum_{i=1,\ldots,9} \beta_i \underbrace{\mathbb{1}(D(c) = 10)}_{\text{Own Income Top Decile}} \times \underbrace{\mathbb{1}(D(\tilde{c}) = i)}_{\text{Parental Income Deciles}} + \varepsilon \tag{A49}$$

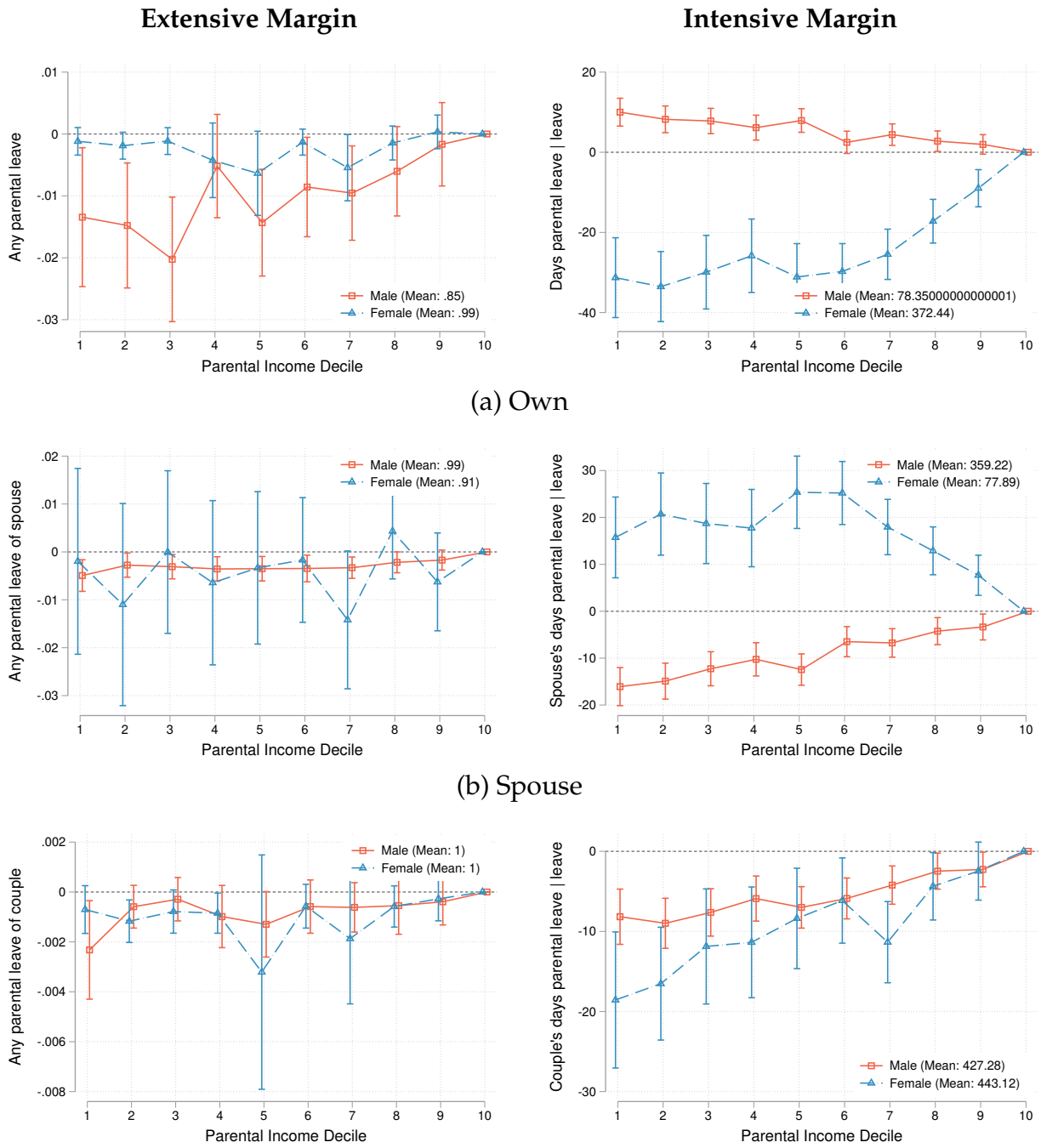
The baseline mean is calculated as the average of the outcome variable among the bottom-decile-income individuals with bottom-decile-income parents. The estimation sample is based on 1,355,750 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. All family outcomes in this figure are conditional on having at least one child. Own lifetime income c is the sum of labor incomes between ages 25 and 45. Parental lifetime income $\tilde{c}$ is the sum of labor, business, and capital income between the individual's ages 0 and 20.

Figure C2: Age at first birth, non-parametric analysis



Note: This figure illustrates the relative difference in the age at first birth of individuals reaching the top-decile of lifetime income with parental income in the bottom decile ($\{\beta_1\}$ of Equation 11). The estimation sample is based on 1,589,961 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. We restrict the sample to individuals with at least one child.

Figure C3: Parental Leave per Child

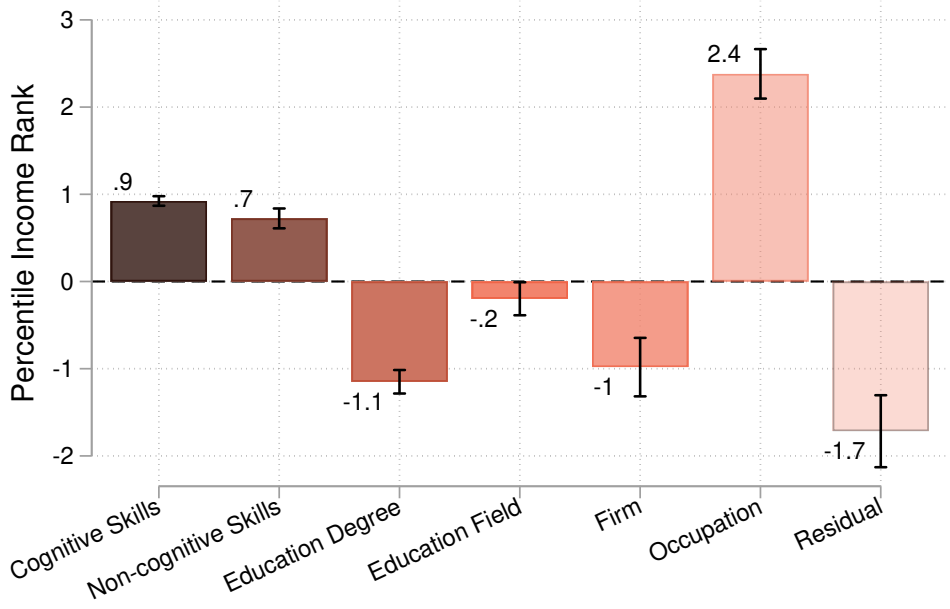


Note: This figure illustrates the difference in parental leave between individuals at the top vs. the bottom of lifetime income by parental income deciles, using top-decile parents as the reference group. It plots the coefficients of the interaction terms, $\{\beta_i\}$, in the regression of family outcome k on a gender-specific non-parametric functions of their own lifetime income c and their parental lifetime income $\tilde{c}$:

$$k = f(c) + g(\tilde{c}) + \sum_{i=1,\ldots,9} \beta_i \underbrace{\mathbb{1}(D(c) = 10)}_{\text{Own Income Top Decile}} \times \underbrace{\mathbb{1}(D(\tilde{c}) = i)}_{\text{Parental Income Deciles}} + \varepsilon \tag{A51}$$

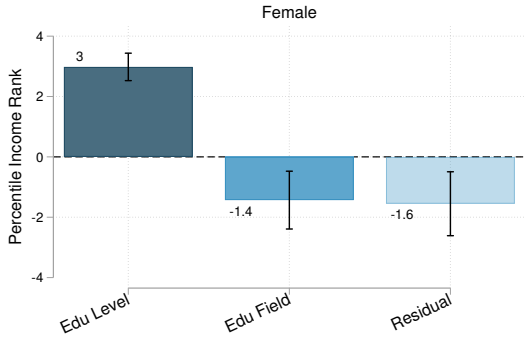
Parental leave is measured individually (Panel (a)), per spouse of the individual (Panel (b)), and jointly by both parents (Panel (c)). The first column shows the extensive margin of taking any parental leave. The second column shows the intensive margin of parental leave days conditional on taking leave. We measure *Parental leave per child* as the total number of days on parental leave per child. We identify the spouse as the other biological parent of the respective child. All panels are limited to individuals born in Sweden with at least one child and the birth cohorts of 1969-1975 for whom we observe parental leave days by age 45 (435,524 individuals). The baseline mean is calculated as the average of the outcome variable among the bottom-decile-income individuals with bottom-decile-income parents. Own lifetime income c is the sum of labor incomes between ages 25 and 45. Parental lifetime income $\tilde{c}$ is the sum of labor, business, and capital income between the individual's ages 0 and 20.

Figure C4: Labor market and human capital inequality

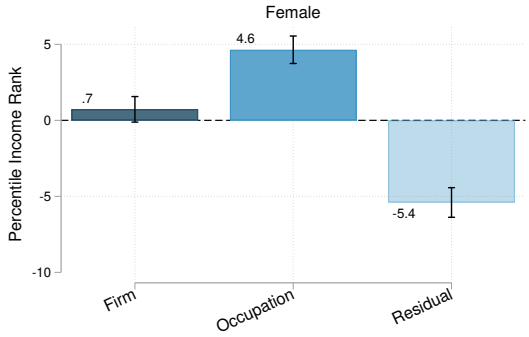


Note: This figure illustrates the relative differences in human capital and labor market outcomes between individuals reaching the top decile of lifetime income from different parental income deciles. It plots the coefficients $\{\beta_1\}$ from Regression 11, where the outcomes are four measures of human capital (cognitive skills, noncognitive skills, education level, and education field) and two aspects of the labor market (firms and occupations). Prior to using these measures as outcomes in Regression 11, we harmonize them to a common scale by mapping each to predicted lifetime-income ranks. Specifically, we estimate a separate Mincerian regressions of lifetime income rank on cognitive and noncognitive skills, education level and field, and fixed effects for firm and occupation. This procedure decomposes lifetime income into human capital and labor market components. We restrict this exercise to men because it includes skill measures from the military draft (560,167 individuals). We report the corresponding estimates for females, omitting the skill measures, in Appendix Figures C5c. Cognitive skill measures are based on test scores for (i) inductive skills, (ii) verbal comprehension, (iii) spatial ability, and (iv) technical understanding. Noncognitive skill measures stem from an interview with a certified psychologist and include scores for (i) emotional stability, (ii) intensity, (iii) psychological energy, and (iv) social maturity. We exploit 13 distinct education-level, 336 field-of-study categories, and 123 occupation codes.

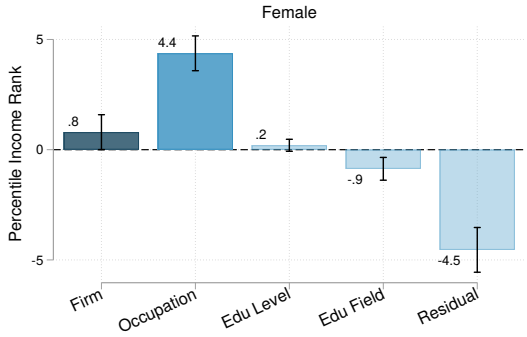
Figure C5: Women: Labor market and human capital inequality



(a) Human Capital Inequality



(b) Labor Market Inequality



(c) Labor Market Inequality conditional on Human Capital

Note: This figure reports the corresponding results for females, using the same approach as Figure 8. Cognitive and non-cognitive skill measures are omitted, as these are only available for men in the military draft data. This figure illustrates the relative difference in human capital and labor market outcomes between individuals reaching the top decile of lifetime income from different parental income deciles. It plots the coefficient $\{\beta_1\}$ in Regression 11 where outcomes are two separate measures of human capital and two separate measures of the labor market. The measures in each figure are harmonized using their predicted lifetime income ranks. Specifically, we use the predicted values of each measure from a Mincerian regression of lifetime income rank on human capital measures (Panel C5a), on labor market measures (Panel C5b), and the combination of the two (Panel C5c). We exploit 13 distinct categories of highest completed education level, 336 categories of field of study, and 123 occupation categories.

Figure C6: Evolution of Cost of Mobility over Time

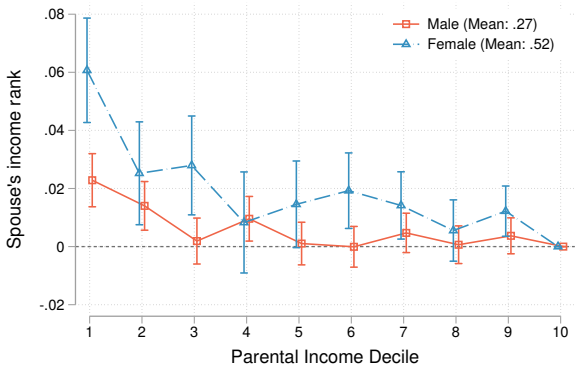


Note: This figure shows the coefficient estimates of running the following regression separately by birth cohort groups:

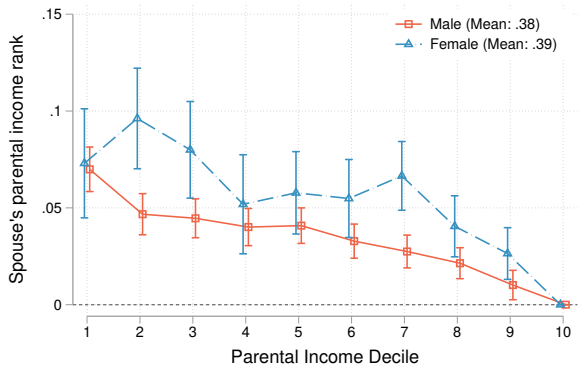
$$k = f(c) + g(\widetilde{c}) + \sum_{i=1,\ldots,9} \beta_i \underbrace{\mathbb{1}(D(c) = 10)}_{\text{Own Income Top Decile}} \times \underbrace{\mathbb{1}(D(\widetilde{c}) = i)}_{\text{Parental Income Deciles}} + \varepsilon \tag{A53}$$

While we interact the control variables $f()$ and $g()$ by gender, we run the interaction term of interest pooled by gender. We categorize individuals into two groups based on their year of birth (YOB): Those born between 1955 and 1965 (black line) and those born between 1965 and 1975 (green line). The estimation sample is 1,589,961 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. Own lifetime income c is measured as the sum of all labor income between ages 25 and 45. Parental lifetime income $\widetilde{c}$ is measured as the sum of the father's and mother's total income over the individual's ages 0 to 20. The number of children includes all biological offspring born before age 45. For more details see Section 2.

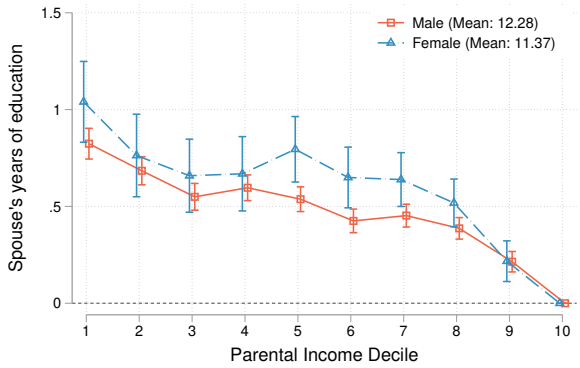
Figure C7: Characteristics of the Spouse



(a) Income Rank Spouse



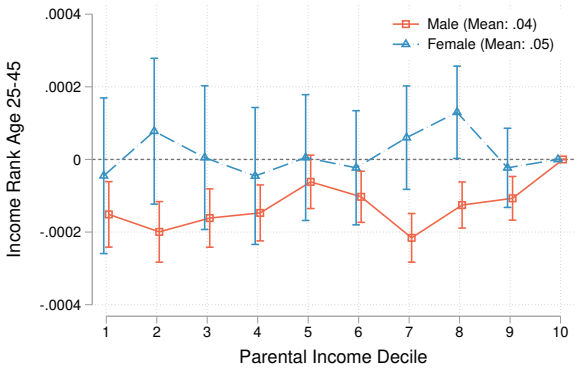
(b) Parental Income Rank Spouse



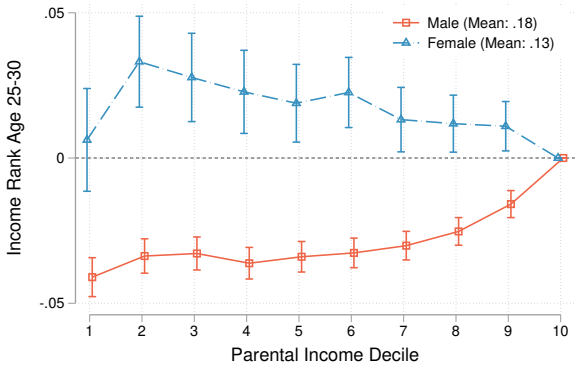
(c) Years of education Spouse

Note: This figure plots the coefficients of our main specification when regressing characteristics of individual i's spouse on their own lifetime income, parental (father + mother) lifetime income, and the interaction (Regression 11). The estimation sample is 822,907 spouses to individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. The estimation sample is smaller than the main sample as we condition on having a spouse at the age of 45. We can identify spouses if they are married or cohabitating with a joint child. We define the income rank of the spouse as the sum of all labor income between ages 25 and 45, ranked in the population of Sweden (Panel (C7a)). Panel (C7b) displays the parental income rank of the spouse, also ranked in the population of Sweden, while Panel (C7c) shows the spouse's years of education.

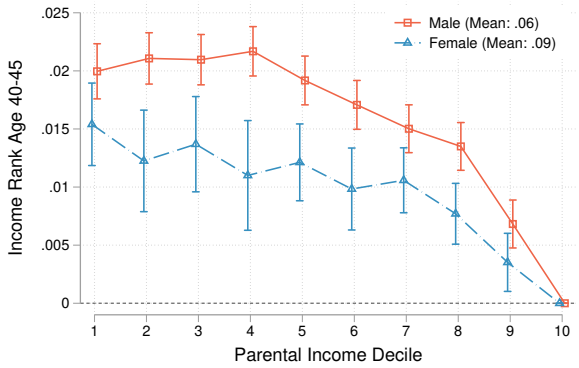
Figure C8: Career dynamics



(a) Income Rank between age 25 and 45



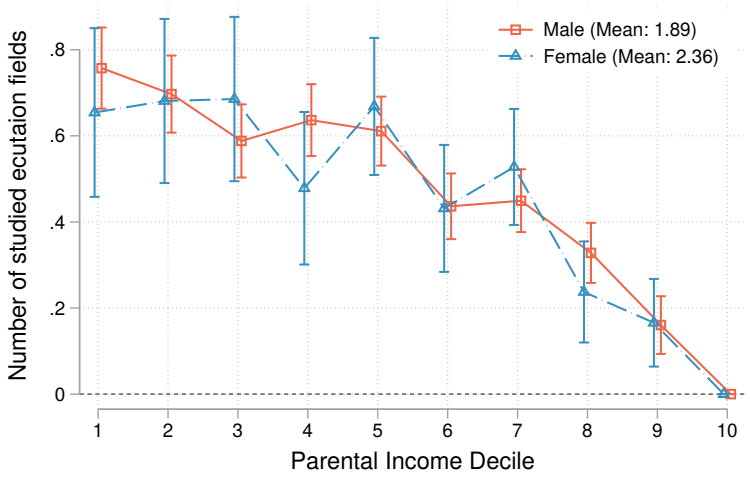
(b) Early Career Income Rank



(c) Late Career Income Rank

Note: This figure displays differences in the income rank position of individuals reaching the top-decile of lifetime income as a function of their parental income deciles ($\{\beta_l\}$ of Equation 11). The outcome of Panel C8a plots the income rank over the individual's career between age 25 and 45. Panel C8b describes the rank early in the career, between age 25 and 30. Panel C8c the rank later in the career, between age 40 and 45. The estimation sample is based on 1,589,961 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. The baseline mean reports the average of dependent variables among the bottom-decile-earning individuals with bottom-decile-earning parents. For more details, see Section 2.

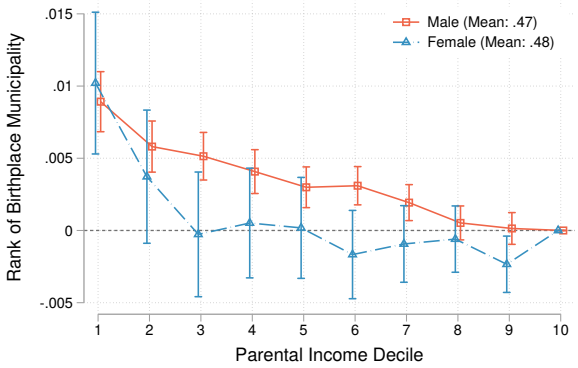
Figure C9: Trial and Error in Education Paths



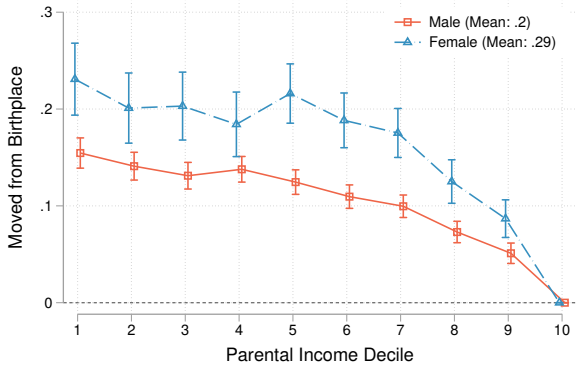
(a) Interaction, β_l

Note: This figure displays the difference in attained study fields of individuals reaching the top-decile of lifetime income as a function of parental income deciles (see Regression 11). For a sample of individuals born between 1970, we define the *number of studies education fields* as the count of fields studied independently of graduating from that field.

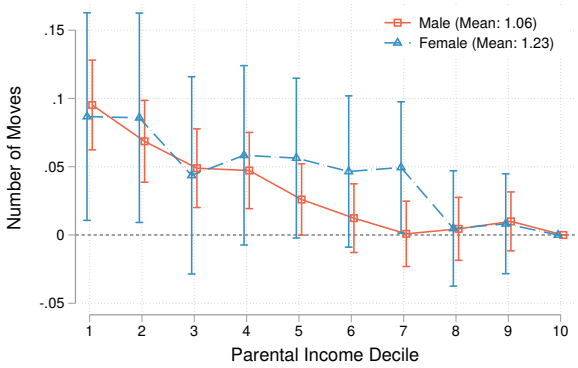
Figure C10: Geographical Mobility



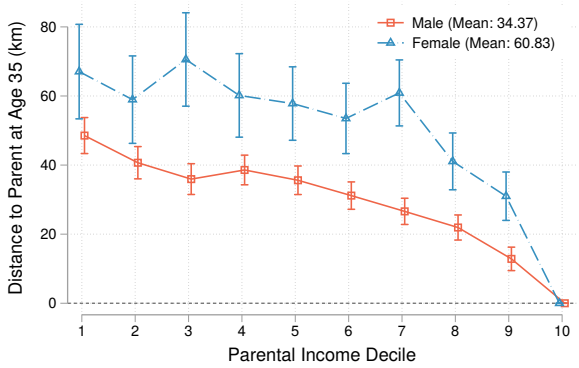
(a) Birthplace Income Rank



(b) Moved From Birthplace



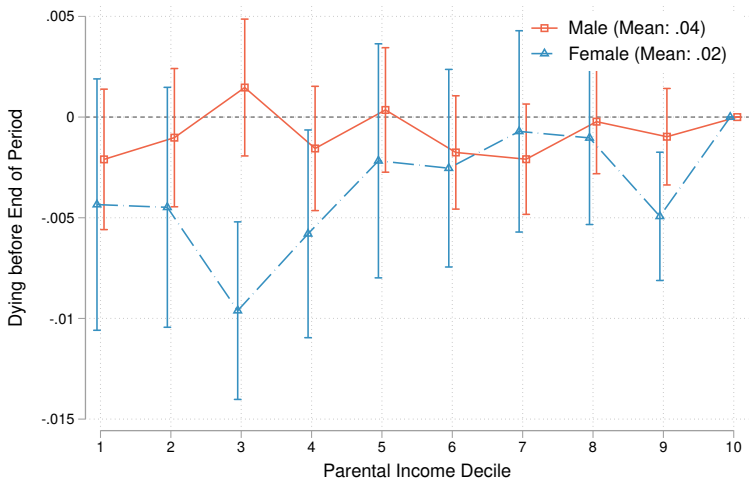
(c) Number of Moves



(d) Distance to Parent

Note: This figure illustrates the relative different geographical mobility of individuals reaching the top-decile of lifetime income as a function of their parental income deciles ($\{\beta_l\}$ of Equation 12). The outcome of Panel C10a measures the average population income rank of working-age individuals living in the municipality at the year of birth. Panel C10b describes an indicator that takes the value one if the individual's county of residence at age 45 differs from the county of birth. In Panel C10c, we measure the dependent variable by counting the number of residence changes between the ages of 35 to 45. Finally, Panel C10d reports the straight-line distance in km between the individual's residence address and a parent's residence address. We use the minimum distance when parents have different residences. The estimation sample is based on 1,589,961 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. The baseline mean reports the average of dependent variables among the bottom-decile-earning individuals with bottom-decile-earning parents. For more details, see Section 2.

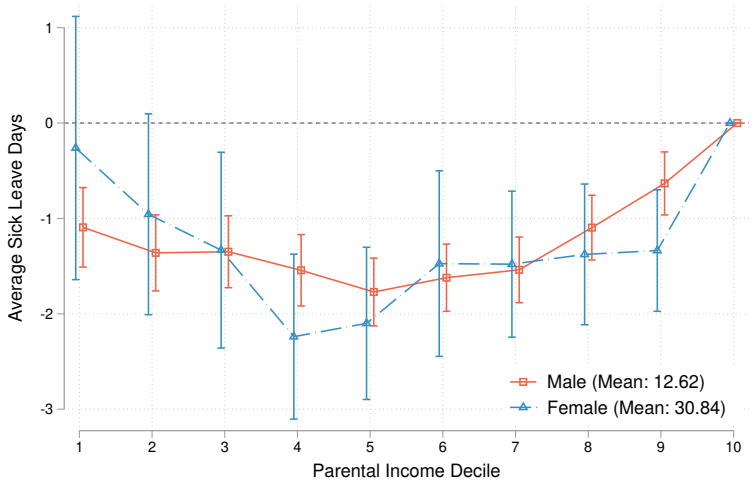
Figure C11: Mortality



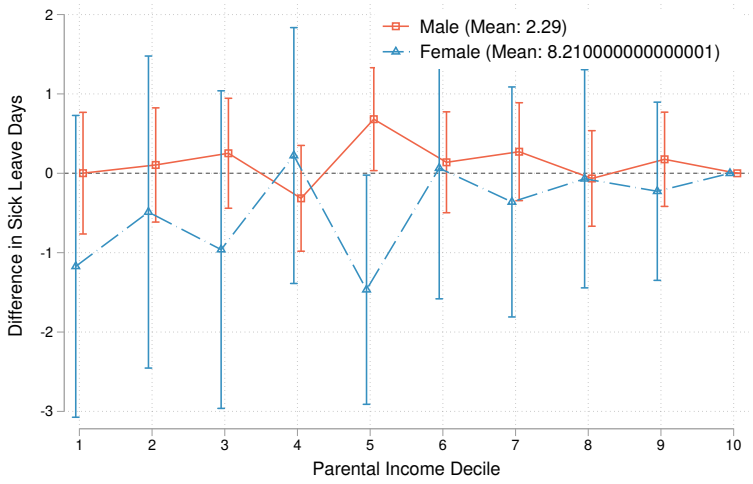
(a) Interaction, β_l

Note: This figure illustrates the relative difference in mortality between individuals reaching the top-decile of lifetime income and individuals in the bottom 90 percent separately by parental income deciles. We use parents in the top income decile as the reference group. Formally, it plots the coefficients of the interaction terms, $\{\beta_l\}$, in the regression of individuals' outcome k on their own lifetime income c and their parental (father + mother) lifetime income $\tilde{c}$ where we interact the regression by gender. We additionally add controls of year and month of birth to the regression. The outcome variable, mortality, is defined as a indicator function that takes the value one if the individual dies at an age before the last observed age of the respective birth cohort. For example, for the cohort born in the year 1970 this would imply that the individuals dies at an age younger than age 50 which is the last point in time we observe the 1970-cohort before the data ends in year 2020. Own lifetime income is measured as the sum of all labor income between ages 25 and 45. Parental lifetime income is measured as the sum of the father's and mother's total income over the individual's ages 0 to 20. The estimation sample is based on 1,589,961 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. We report the mean for individuals in the bottom-decile income bracket with parents in the bottom-decile income bracket. For more details see Section 2.

Figure C12: Days of Sick Leave



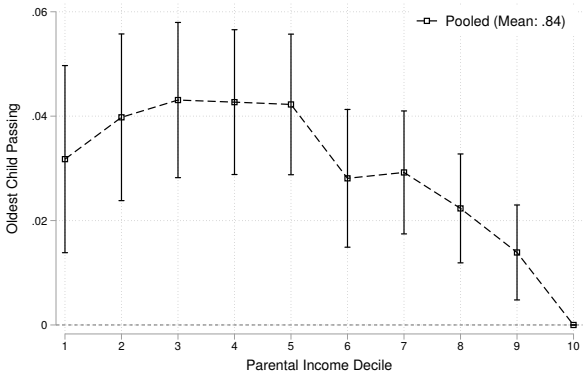
(a) Level



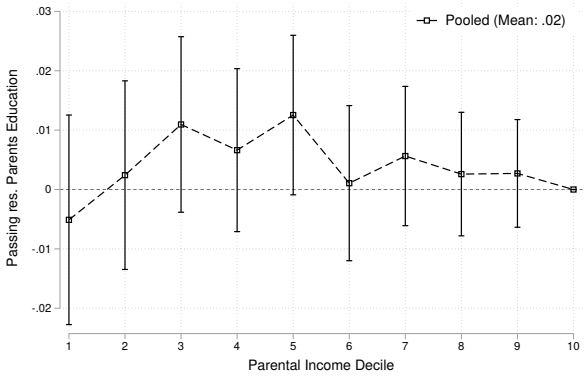
(b) Change

Note: This figure illustrates the relative difference in sick leave days between individuals reaching the top-decile of lifetime income and individuals in the bottom 90 percent separately by parental income deciles. We use parents in the top income decile as the reference group. Formally, it plots the coefficients of the interaction terms, $\{\beta_l\}$, in the regression of outcome k on individuals' own lifetime income c and their parental (father + mother) lifetime income $\widetilde{c}$, where we interact the regression by gender. We additionally control for year of birth and month in the regression. Due to data limitations, we compute the average number of sick leave days per year between age 35 and 45. The outcome variable, *difference in sick leave days*, is defined as the change in the average number of days on sick leave per year measured between the ages of 40 and 45 and the the ages of 35 and 40. We exclude sick leaves related to pregnancy. Own lifetime income is measured as the sum of all labor income between ages 25 and 45. Parental lifetime income is measured as the sum of the father's and mother's total income over the individual's ages 0 to 20. We rank these income measures within the birth cohort. The estimation sample is based on 1,589,961 individuals born in Sweden between 1955 and 1975 from parents who were also born in Sweden. Due to data limitations, we can measure the days of sick leave for the cohorts born in 1959 to 1975. We report the mean for individuals in the bottom-decile income bracket with parents in the bottom-decile income bracket. For more details see Section 2.

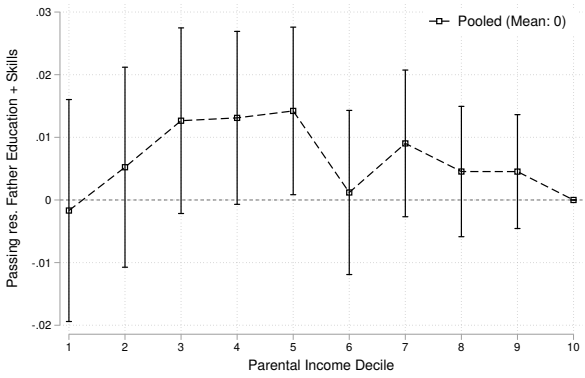
Figure C13: Children's Educational Achievements



(a) Children's Education



(b) Children's Education, controlling for parental education



(c) Children's Education, controlling for parental education & skills

Note: This figure illustrates the difference in school performance of the oldest child between individuals reaching the top-decile of lifetime income and individuals in the bottom 90 percent by parental income deciles, relative to the same difference for those with parents at top decile. Specifically, it plots the coefficients of the interaction terms, $\{\beta_i\}$, in specification 11. The outcome variable, the school performance of the oldest child, is an indicator of whether the individual's oldest child passes English, Swedish, and Mathematics on the grade-9 national exam. In the regression, we additionally control for the own's and the child's year of birth. In Panels C13b and Panel C13c, the outcome is first residualized by regressing it on parent's education degree and field. In the last panel, we additionally use the father's cognitive and non-cognitive skills. In this figure, we restrict the sample to individuals who have at least one child and to individuals born between 1963 and 1969 due to data limitations in the national test exams.

D Appendix Tables

Table A.1: 4 digit education field codes

Code	Description (English Translation)
010a	Broad, general education, general orientation
010b	Broad, general education, social science and humanities focus
010c	Broad, general education, natural science focus
010x	Broad, general education, other specialization
080z	Learning to read and write for adults
090z	Personal development
140z	Pedagogy and teacher training, general education
142z	Pedagogy
143a	Teacher training, specialization preschool
143b	Teacher education, focus on leisure activities
143x	Other teacher training for preschool and leisure activities
144a	Teacher training, primary school lower ages, general subject orientation
144b	Teacher training, primary school lower ages, mathematics and nature-oriented subjects
144c	Teacher training, primary school lower ages, Swedish and community-oriented subjects
144x	Other teacher training for the lower ages of primary school
145a	Subject teacher training, mathematics, data and science
145b	Subject teacher training, history, social science and economics
145c	Subject teacher training, language
145d	Subject teacher education, mathematics, data, and natural sciences in combination with history, social sciences and economics
145e	Subject teacher training, mathematics, data, and natural sciences in combination with languages
145 f	Subject teacher training, history, social science and economics in combination with languages
145g	Subject teacher training, mathematics, data and science in combination practical-aesthetic subjects
145h	Subject teacher training, history, social science and economics in combination with practical-aesthetic subjects
145i	Subject teacher training, language, in combination with practical-aesthetic subjects
145x	Other subject teacher training
146a	Teacher training, industry and crafts
146b	Teacher training, trade and administration
146c	Teacher training, hotel and restaurant
146d	Teacher training, food and confectionery
146th	Teacher training, natural farming
146 f	Teacher training, care and social care
146g	Teacher training, media production
146h	Teacher education, music
146i	Teacher training, visual and visual arts
146j	Teacher training, dance and theatre
146k	Teacher training, wood and metal work and textile work
146l	Teacher training, sports and health
146m	Teacher training, household science
146x	Teacher training in other vocational subjects and practical-aesthetic subjects
149a	Special education teacher training
149b	Home language and mother tongue teacher training
149c	School leader training
149x	Other education in pedagogy and teacher training
210z	Art and media, general education
211a	Visual and visual arts education
211b	Art science and art history
211x	Other education in the visual and visual arts
212a	Music
212b	Dance, theater and drama
212c	Music, dance, film and theater science and history
212x	Other education in music, dance and drama
213a	Media production, general education
213b	Film, radio and TV production

Continued on next page

Table A.1 – continued from previous page

Code	Description (English Translation)
213c	Graphic technology and bookbinding
213d	Illustration, advertising, graphic design and photography
213x	Other education in media production
214a	Fashion design
214b	Interior design
214c	Industrial and product design
214x	Other education in design
215z	Crafts
219z	Art and media, other and unspecified education
220z	Humanities, general education
221a	Theology bachelor's degree and equivalent
221b	Pastor, missionary and deacon training
221x	Other education in religion
222a	General language skills and general language education
222b	English
222c	German
222d	French
222e	Spanish
222 f	Russian
222g	Swedish as a foreign language
222x	Other education in foreign languages
223a	Swedish and Nordic languages
223b	Other mother tongues and sign languages
223c	Literary studies and literary history
223x	Other education in Swedish and literary studies
225a	History
225b	Archaeology
225c	Cultural preservation, cultural administration and museum science
225x	Other education in history and archaeology
226z	Philosophy and logic
229z	Humanities, other and unspecified education
310a	Behavioral Science, General Education
310b	Social science, general education
310x	Other broad education in social and behavioral sciences
311a	Psychology education
311b	Psychology training, including completed internship (PTP)
311c	Psychotherapist training
311x	Other education in psychology
312a	Sociology
312b	Social anthropology and ethnology
312c	Cultural and social geography
312x	Other education in sociology, ethnology and cultural geography
313z	Political science
314z	Economics and economic history
319z	Social and behavioral sciences, other and unspecified education
320z	Journalism and information, general education
321a	Journalism
321b	Media and communication science
321x	Other education in journalism and media science
322a	Librarian education and library and information science
322b	Documentation and information science
322x	Other education in library and documentation science
329z	Journalism and information, other and unspecified education
340a	Economics education and business economics
340b	Business and Administration
340x	Other broad education in business administration, trade and administration
341z	Purchasing, sales and distribution
342z	Marketing
343z	Banking, insurance and financing

Continued on next page

Table A.1 – continued from previous page

Code	Description (English Translation)
344z	Accounting and Taxation
345a	Leadership, organization and governance
345b	Administration and management
345c	Personnel management
345x	Other training in management and administration
346a	Medical secretarial training
346x	Other training in office services and secretarial services
347z	Workplace-related training
349z	Business administration, trade, administration, other and unspecified education
380a	Legal education (Bachelor of Laws and equivalent)
380b	Legal education including completed legal qualification
380x	Other education in law and jurisprudence
420z	Biology and environmental science, general education
421a	Biology
421b	Biochemistry, toxicology, pharmacology and nutrition
421x	Other education in biology and biochemistry
422z	Environmental science
429z	Biology and environmental science, other and unspecified education
440z	Physics, chemistry and earth science, general education
441z	Physics
442z	Chemistry
443z	Earth science and natural geography
449z	Physics, chemistry and geoscience, other and unspecified education
460z	Mathematics and science, general education
461z	Mathematics
462z	Statistics
469z	Mathematics and science, other and unspecified education
480z	Data, general education
481a	Systems science and software engineering
481b	Computer science and computer science
481x	Other education in computer science and systems science
482z	Computing
489z	Data, other and unspecified education
520a	Civil engineering education, general technology and technical physics
520b	Engineering education, general technology and technical physics
520x	Other broad education in technology and technical industry
521a	Civil engineering education, majoring in mechanical engineering
521b	Engineering education, majoring in mechanical engineering
521c	Industry and workshop technology
521x	Other education in mechanical engineering and workshop technology
522a	Master's degree in engineering, majoring in energy and electrical engineering
522b	Engineering education, specialization energy and electrical engineering
522c	Electrical installation and electrical mechanics
522d	Energy, operation and maintenance
522x	Other education in energy and electrical engineering
523a	Master's degree in engineering, majoring in electronics, computer technology and automation
523b	Engineering education, majoring in electronics, computer technology and automation
523c	Electronics, telecommunications and computer engineering work
523d	Automation, control and regulation technology
523x	Other education in electronics, computer technology and automation
524a	Civil engineering education, specialization chemical and biotechnology
524b	Engineering education, specialization chemical and biotechnology
524c	Laboratory technology
524d	Process industry technology
524x	Other education in chemical and biotechnology
525a	Master's degree in engineering, specialization vehicle and vehicle technology
525b	Engineering education, focus on vehicle and craft technology
525c	Vehicle technology
525d	Naval and aeronautical engineering

Continued on next page

Table A.1 – continued from previous page

Code	Description (English Translation)
525x	Other education in vehicle and craft technology
526a	Civil engineering education, majoring in industrial economics and organization
526b	Engineering education, majoring in industrial economics and organization
526x	Other education in industrial economics and organization
529a	Civil engineering education, other and unspecified education
529b	Engineering education, other and unspecified education
529x	Other education in technology and technical industry
540a	Master's degree in engineering, majoring in materials technology
540b	Engineering education, specialization materials technology
540x	Other broad training in materials and manufacturing
541a	Master's degree in engineering, food specialization
541b	Engineering education, focus on food
541c	Food manufacturing and handling, general education
541d	Meat and charcuterie
541e	Bakery and patisserie
541x	Other training in food manufacturing and handling
542a	Civil engineering education, focus on textile technology
542b	Engineering education, focus on textile technology
542c	Manufacture of textiles and clothing
542d	Shoe and leather work
542x	Other education in textiles, clothing and leather
543a	Civil engineering education, specialization wood, paper, glass, porcelain and plastic
543b	Engineering education, focus on wood, paper, glass, porcelain and plastic
543c	Wood technical work
543d	Pulp and paper manufacturing
543e	Glass and porcelain manufacturing
543 f	Manufacture of plastic and rubber products
543x	Other training in manufacturing in specific materials
544a	Civil engineering education, focus on rock and mineral technology
544b	Engineering education, focus on rock and mineral technology
544c	Mining and iron mill work
544x	Other education in rock and mineral engineering
549z	Materials and manufacturing, other and unspecified training
580a	Civil engineering education, broad focus on construction, i.a. land surveying
580b	Engineering education, broad focus on construction, i.a. land surveying
580x	Other broad education in civil engineering and building technology
581a	Master's degree in civil engineering, majoring in civil engineering and social environment
581b	Engineering education, focus on civil engineering and social environment
581c	Architectural education
581d	Landscape architect and landscape engineer education
581x	Other education in civil engineering and architecture
582a	Civil engineering education, majoring in construction technology
582b	Engineering education, majoring in construction technology
582c	Construction work, general education
582d	Building carpentry
582e	Painting
582 f	Plumbing training
582g	Concrete, construction and road work
582 h	Masonry education
582 i	Sheet metal warehouse
582x	Other education in building and civil engineering
589z	Civil engineering and building technology, other and unspecified education
620z	Agriculture, gardening, forestry and fishing, general education
621a	Agronomist training
621b	Landmaster training
621c	Education in agriculture, general
621d	Education in agriculture and horticulture
621e	Education in animal husbandry
621 f	Training in horse husbandry and hippologist training

Continued on next page

Table A.1 – continued from previous page

Code	Description (English Translation)
621g	Education in animal care
621x	Other education in agriculture
622a	Horticulture training
622b	Horticultural engineering and horticultural technician training
622c	Training in horticulture
622x	Other education in gardening
623a	Master hunter and forest scientist training
623b	Forester, forest technician and forest engineer training
623c	Education in forestry
623x	Other training in forestry
624z	Fishing and aquaculture
629z	Agriculture, gardening, forestry and fishing, other and unspecified education
640a	Veterinary education
640x	Other training in animal health care
720z	Health care, general education
721a	Medical education
721b	Medical education, including completed public service (AT)
721c	Medical training, specialist competence in surgery, anesthesia and intensive care
721d	Medical education, specialist competence in internal medicine
721e	Medical training, specialist competence in children's medicine
721 f	Medical education, specialist competence in geriatrics
721g	Medical education, specialist competence in general medicine
721 h	Medical training, specialist competence in psychiatry
721 i	Medical education, specialist competence in radiology
721j	Medical education, clinical laboratory specialties
721k	Medical education, other specialist skills
721x	Other education in medicine
723a	Nursing education, basic education
723b	Nursing education, general health care
723c	Nursing training, anaesthesia, intensive care, emergency, accidents, surgery
723d	Nursing education, children and young people
723e	Nursing education, geriatric care
723 f	Nursing education, psychiatric care
723g	Nursing education, district nursing activities
723 h	Nursing education, midwifery
723i	Nursing education, diagnostic radiology and oncology
723j	Nursing education, other specialization
723m	Other nursing education, general care and social care
723n	Other nursing education, focusing on healthcare
723 o	Other nursing education, focusing on psychiatric care
723p	Other nursing education, focusing on elderly care
723q	Other nursing education, focusing on the mentally retarded
723r	Other nursing education, focusing on the disabled
723p	Other nursing education, other orientations
723x	Other education in nursing
724a	Dental education
724b	Dental education, including specialization
724c	Dental hygienist training
724d	Dental nursing education
724e	Dental technician training
724x	Other education in dentistry
725a	Optician training
725b	Audiology training
725c	Orthopedic engineering education
725d	Biomedical analyst training
725e	Hospital physicist training
725 f	X-ray nursing training
725x	Other technically oriented nursing education
726a	Physiotherapy education

Continued on next page

Table A.1 – continued from previous page

Code	Description (English Translation)
726b	Occupational therapist training
726c	Speech therapy training
726d	Dietician training
726x	Other training in therapy, rehabilitation and dietary treatment
727a	Pharmacist education
727b	Recipe training
727x	Other education in pharmacy
729z	Health care, other and unspecified education
760z	Social work and care, general education
761a	Children and youth, general education
761b	Child care
761c	Leisure activities for children and young people
761x	Other education in children and youth
762a	Social service and care, general education
762b	Social care education, elderly and disabled
762c	Social care training, social pedagogical treatment
762d	Sociology training
762e	Sociology education, elderly and disabled
762 f	Sociologist training, social pedagogical treatment
762g	Study and vocational guidance training
762x	Other education in social work and guidance
769z	Social work and care, other and unspecified education
810z	Personal services, general education
811a	Hotels, restaurants and large households, general education
811b	Training for hotels
811c	Training for restaurant
811d	Training for large households
811x	Other training in hotels, restaurants and large households
812z	Tourism, travel and leisure
813z	Sports and wellness
814a	Home and Household Science, General Education
814b	Training for housekeeping
814x	Other training in household services and housekeeping
815a	Training for hair care
815x	Other training in hair and beauty care
819z	Personal services, other and unspecified training
840a	Transport services, general education
840b	Training for land transport
840c	Training for shipping
840d	Training for aviation
840e	Training for postal service
840x	Other training in transport services
850z	Environmental protection and environmental protection, general education
851z	Environmental technology and environmental control
852z	Nature conservation and animal welfare
853z	Cleaning and waste management
859z	Environmental care and environmental protection, other and unspecified education
860z	Security services, general education
861a	Police training
861b	Training for fire protection and other emergency services
861c	Civil defense training
861x	Other training in community safety
862z	Work environment and worker protection
863a	Professional officer training
863b	Reserve officer training
863c	Tactical training for military personnel
863d	Vocational training for military personnel
863e	Staff training for military personnel
863 f	Higher management training for military personnel

Continued on next page

Table A.1 – continued from previous page

Code	Description (English Translation)
863x	Other military training
869z	Security services, other and unspecified training
999z	Unknown

Note: The table reports the descriptions of the 4 digit education field of study codes (336 categories). Field of study classifications were originally provided in Swedish. For the purposes of analysis and presentation, all translations into English were carried out by the authors using ChatGPT. Any misinterpretations or inaccuracies are the authors' responsibility.

Table A.2: 2 digit education degree codes

Code	Description (English Translation)
00	Preschool education
10	Pre-secondary education, shorter than 9 years
20	Pre-secondary education, 9 (10) years
31	Upper secondary education, shorter than 2 years
32	Upper secondary education, 2 years
33	Upper secondary education, 3 years
41	Post-secondary education, shorter than 2 years
52	Post-secondary education, 2 years
53	Post-secondary education, 3 years
54	Post-secondary education, 4 years
55	Post-secondary education, 5 years or longer
60	Other postgraduate (research) education
62	Licentiate degree education
64	Doctoral education
99	Unknown

Note: The table reports the descriptions of the 3 digit education degree codes (47 categories). Education degree classifications were originally provided in Swedish. For the purposes of analysis and presentation, all translations into English were carried out by the authors using ChatGPT. Any misinterpretations or inaccuracies are the authors' responsibility.

Table A.3: 3-digit occupation codes

Code	Description (English Translation)
111	Senior officials and politicians
112	Senior officials in interest organizations
121	Chief executives and managing directors
122	Production and operations managers
123	Managers of specialized functions
131	Managers of small enterprises and units
211	Physicists, chemists, and related professionals
212	Mathematicians and statisticians
213	IT specialists
214	Civil engineers, architects, and related professionals
221	Specialists in biology, agriculture, forestry, etc.
222	Health professionals (except nursing)
223	Midwives; nurses with specialized competence
231	University and higher education teachers
232	Upper secondary school teachers
233	Primary school teachers
234	Special education teachers
235	Other teaching professionals with theoretical specialist competence
241	Economists, marketers, and HR officers
242	Legal professionals
243	Archivists, librarians, and related professionals
244	Social scientists and linguists (except social workers)
245	Journalists, artists, actors, and related professionals
246	Clergy
247	Public administration professionals
248	Administrators in interest organizations

Continued on next page

Table A.3 – continued from previous page

Code	Description (English Translation)
249	Psychologists, social workers, and related professionals
311	Engineers and technicians
312	Computer technicians and operators
313	Photographers; audio/video technicians, hospital techs, etc.
314	Pilots, ship officers, and similar
315	Safety and quality inspectors
321	Agricultural and forestry technicians
322	Physiotherapists, dental hygienists, etc.
323	Nurses
324	Biomedical analysts
331	Preschool teachers and leisure-time pedagogues
332	Other teachers and instructors
341	Salespersons, buyers, brokers, etc.
342	Agents, intermediaries, etc.
343	Accountants, administrative assistants, etc.
344	Customs, tax, and social insurance officials
345	Police officers
346	Treatment assistants, recreation leaders, etc.
347	Illustrators, entertainers, professional athletes, etc.
348	Pastors and other religious workers
411	Office secretaries and data entry clerks
412	Bookkeeping and accounting assistants
413	Warehouse and transport clerks
414	Library assistants and similar
415	Mail carriers and sorters
419	Other office staff
421	Cashiers, tellers, and related clerks
422	Customer service clerks
511	Travel attendants and related workers
512	Institutional kitchen and restaurant staff
513	Care and welfare workers
514	Hairdressers and other personal service workers
515	Security personnel
521	Fashion and other models
522	Retail salespersons and demonstrators
611	Crop growers in agriculture and horticulture
612	Animal breeders and handlers
613	Mixed crop and livestock producers
614	Forestry workers
615	Fishers and hunters
711	Miners, quarry workers, stonecutters
712	Building and construction workers
713	Building trades workers
714	Painters, varnishers, chimney sweeps, etc.
721	Metal casters, welders, sheet metal workers, etc.
722	Blacksmiths, toolmakers, etc.
723	Machine and engine mechanics
724	Electrical, telecom, and electronics repairers
731	Precision mechanics and similar trades
732	Potters, glassworkers, decorative painters, etc.
733	Craftworkers in wood, textiles, leather, etc.
734	Graphic arts and related trades workers
741	Food processing workers
742	Furniture and model carpenters
743	Tailors, cutters, upholsterers, etc.
744	Tanners, leather workers, shoemakers
811	Ore processing operators, well drillers, etc.
812	Operators in steel and metal production
813	Operators in glass and ceramic production

Continued on next page

Table A.3 – continued from previous page

Code	Description (English Translation)
814	Operators in wood and paper industries
815	Operators in the chemical industry
816	Power plant operators and related workers
817	Industrial robot operators
821	Machine operators in metal and mineral processing
822	Machine operators in chemical industries
823	Rubber and plastic product machine operators
824	Machine operators in the wood industry
825	Machine operators in printing and paper industries
826	Textile, fur, and leather machine operators
827	Food product machine operators
828	Assemblers
829	Other machine operators and assemblers
831	Locomotive drivers and related workers
832	Motor vehicle drivers
833	Agricultural and mobile machinery operators
834	Deck crew and related ship workers
911	Street and market vendors
912	Cleaners and helpers
913	Kitchen and restaurant assistants
914	Newspaper distributors, caretakers, etc.
915	Sanitation and recycling workers
919	Other service workers
921	Helpers in agriculture, gardening, forestry, and fishing
931	Construction laborers
932	Manual packers and other factory workers
933	Freight handlers and couriers
011	Armed forces
	Unknown categories

Note: The table reports the descriptions of the 3-digit occupations. Occupational classifications were originally provided in Swedish. For the purposes of analysis and presentation, all translations into English were carried out by the authors using ChatGPT. Any misinterpretations or inaccuracies are the authors' responsibility.