



Genetic Propensity for Education in Labor Market and Health Trajectories across the Working Life

Stefano Lombardi¹ **Nurfatima Jandarova**² Kristina Zguro³ Jarkko Harju^{2, 1} Aldo Rustichini⁴ Andrea Ganna^{3,5}

¹VATT Institute for Economic Research, Helsinki, ²FIT, Tampere University, Finland ³Institute for Molecular Medicine Finland, ⁴University of Minnesota, USA ⁵Broad Institute of MIT and Harvard, USA

Genetics Friday

Centre for Fertility and Health, Norwegian Institute of Public Health

October 9, 2026



Origins and persistence of socioeconomic inequality

- **Choice and luck** (“accident of birth”) (Cunha and Heckman 2007)
- Genetics and environment interact to shape individual outcomes
- **Polygenic indices (PGIs)** summarise genetic predispositions
- **Existing evidence** on association between PGIs and income remains limited
 - *static* estimates that do not capture income accumulation over the life cycle
 - *coarse*, self-reported measures of income
 - *limited* evidence on key *mediating channels* such as employer and occupational sorting



This paper

Research questions

- How the genetic endowment influences individual outcomes over the **life cycle**?
- Role of **firms** in mediating genetic gradients in income trajectories

What we do

- Link Finnish matched employee-employer registers with genotype data
- Follow graduates annually from graduation up to 25 years later
- Analyse patterns in labour income, firm sorting and health trajectories by PGI



Contributions

Sociogenomic literature use cross-section with coarse self-reported income: Carvalho (2025), Ghirardi et al. (2024), Rustichini et al. (2023), Barth et al. (2020), Rimpler et al. (2018)

Distinguish inequality at entry vs divergent career growth

Wage dispersion and worker-firm sorting use estimates of latent worker “skill”: Card et al. (2018), Song et al. (2019), Kline (2024)

Analyse job mobility and firm sorting as mediators of genetic gaps



Preview of results

Favourable genetic endowment (higher PGI for education)

- is not associated with income level differences at graduation;
- predicts **steeper income trajectory**;
- acts through indirect genetic background of **parents (fathers)**;
- manifests only among **tertiary-educated**;
- contributes to steeper income path thanks to **firm mobility**;
- is weakly associated with health indices



Data

Genotype data



176 523 genotyped individuals from Finnish biobanks ($\mathbf{G}_i$)

Polygenic index for years of education (EA-PGI)

- weighted sum of genotype vector EA-PGI = $\mathbf{G}_i \hat{\beta}$
- measures predisposition to education (including skills and other traits) ► Density
- $\hat{\beta}$ (out-of-sample): largest GWAS of educational attainment (Okbay et al. 2022)



Annual registries 1987-2019

Full population coverage

- **basic records:** gender, age, birth year
- **education records:** highest level, graduation year, field and institution ID
- **matched employee-employer** structure: firm ID, occupation, industry
- **income records:** labour income before tax
- **healthcare records:** we construct Charlson Comorbidity Index



Analysis sample

	Person-year	Person
Panel of genotyped individuals	5 374 521	176 523
Keep graduates only	3 215 453	98 810
Keep graduates with non-missing graduation year	3 215 453	98 810
Graduated between 1970 and 2020	3 139 889	96 186
Observed between 0 and 25 years since graduation	1 692 473	96 166
Followed from 0 years since graduation	1 000 872	57 956
Followed at least up to age 30 (if secondary)	963 715	51 056

Construct **sample weights** based on full-population data [► Balance table](#)



Results



Trajectory estimation

$$y_{icmt} = \alpha + \tau_c + \tau_m + \beta_t PGI_i + \gamma X_i + \varepsilon_{icmt}$$

y_{icmt} outcome of person i born in year c observed in year m at t years since graduation

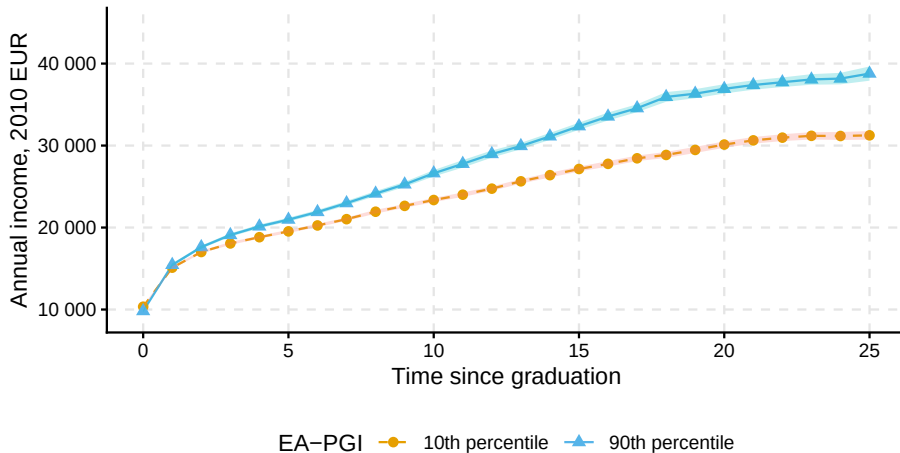
PGI_i standardised EA-PGI

X_i covariates (gender, first ten genetic PCs and biobank indicator)

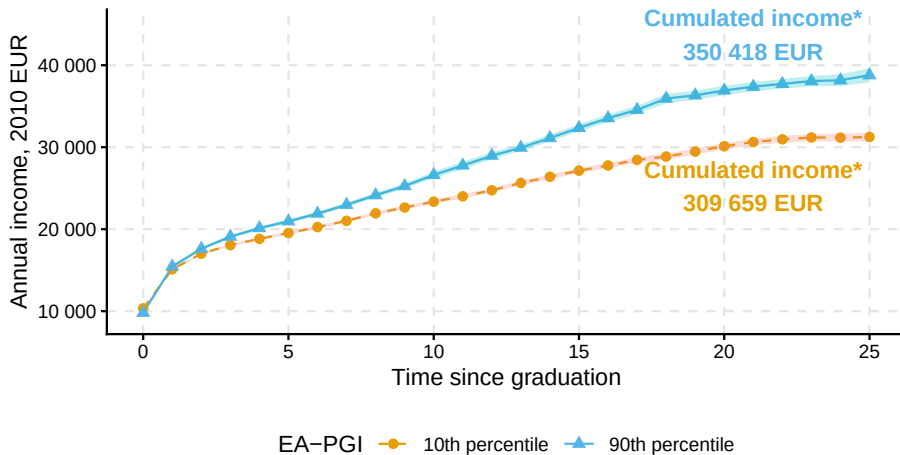
β_t coefficient of own genetics (+ other environmental factors)

Baseline analysis: average $\hat{y}_t$ at 10th and 90th percentiles of EA-PGI [► Deciles](#)

EA-PGI predicts income trajectory



EA-PGI predicts income trajectory



* discounted at 3%



Trajectory estimation with parental EA-PGI

Using 12 871 family trios (genotyped and imputed)

$$y_{icmt} = \alpha + \tau_c + \tau_m + \beta_t PGI_i + \delta_t^m PGI_i^m + \delta_t^f PGI_i^f + \gamma X_i + \varepsilon_{icmt}$$

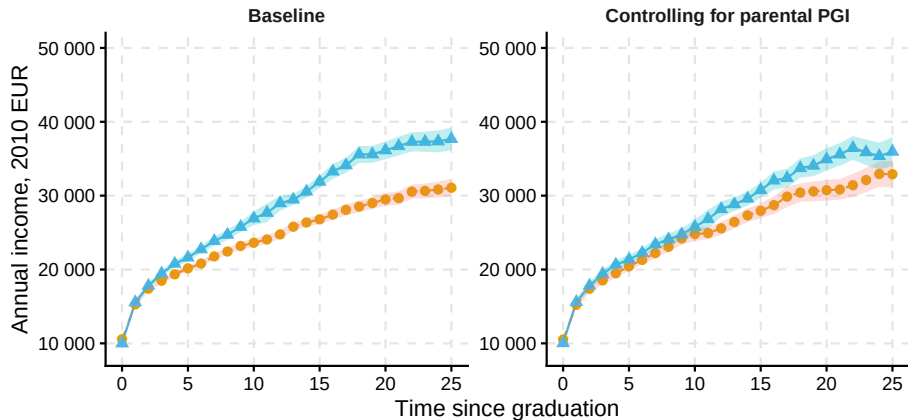
β_t captures direct association with genetic endowment

δ_t^m and δ_t^f reflect both indirect association via parents' genes and family environment

► Baseline

► Replication with yedu

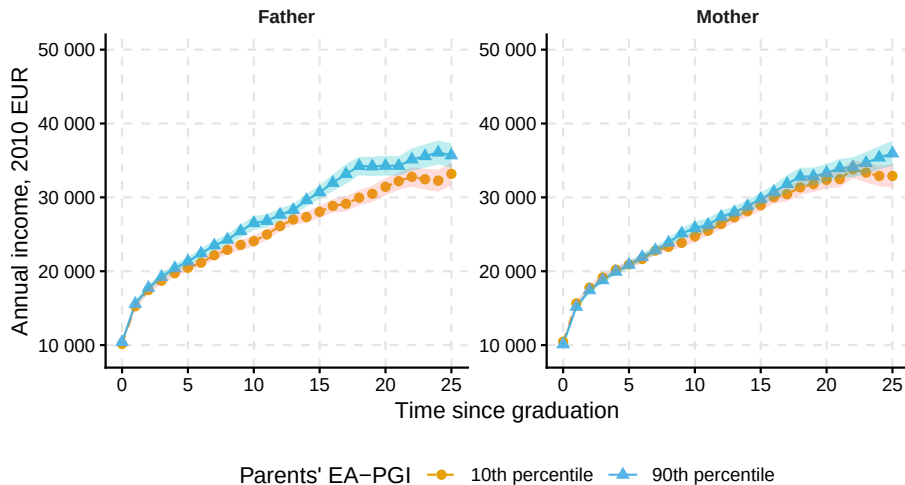
Income gap by EA-PGI shrinks by 53% in family analysis



◀ Weighted

EA-PGI —●— 10th percentile —▲— 90th percentile

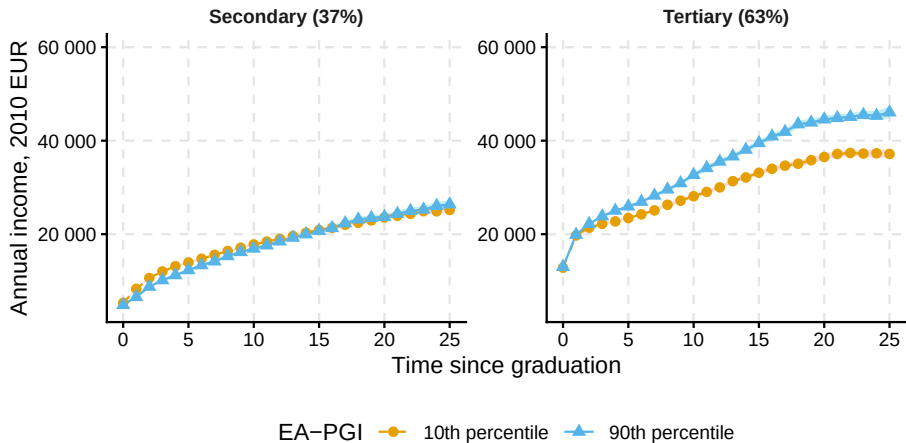
Father EA-PGI predicts children's income trajectories





Mechanisms: Education

Income trajectory gap only among tertiary-educated



AKM decomposition

Using full population registry, estimate

$$y_{it} = \mathbf{X}_{it}\beta + \psi_{J(i,t)} + \theta_i + \varepsilon_{it}$$

y_{it} monthly labour income of worker i in year t

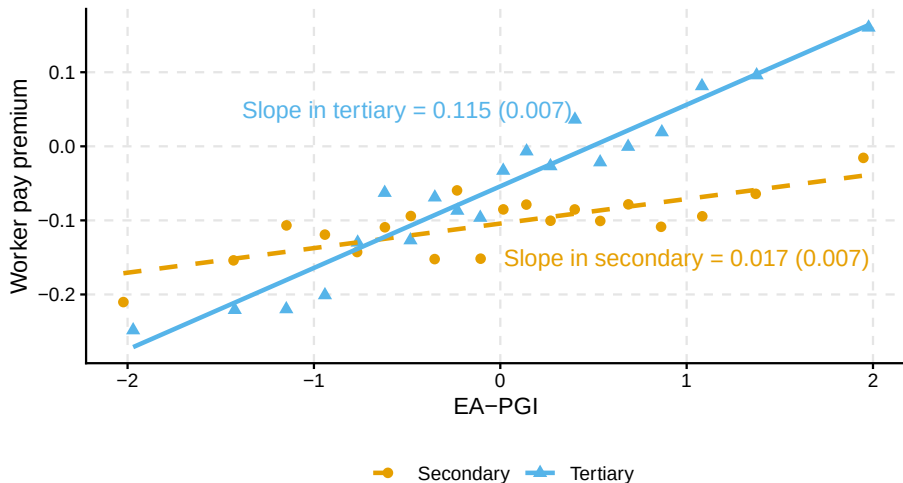
$\mathbf{X}_{it}$ education fully interacted with calendar year and cubic age polynomial

$\psi_{J(i,t)}$ firm fixed effect (*firm pay premium*)

θ_i worker fixed effect (*worker pay premium*)

► AKM summary

EA-PGI associated with worker pay premium among tertiary-educated



Worker pay premium, selection and returns to education

	Dependent variable: Worker productivity index			
	(1)	(2)	(3)	(4)
Secondary education $\times$ EA-PGI	0.017* (0.007)	0.010 (0.007)	0.011 (0.007)	0.007 (0.007)
Tertiary education $\times$ EA-PGI	0.115*** (0.007)	0.086*** (0.007)	0.082*** (0.007)	0.016* (0.006)
Level	Yes	Yes	Yes	Yes
Field	No	Yes	Yes	Yes
Level $\times$ Field	No	No	Yes	No
Institution ID	No	No	No	Yes
Obs.	31,866	31,709	31,702	31,574
Avg. obs. per cell	15,933	352	145	23
Adj R ²	0.198	0.251	0.256	0.330
RMSE	0.836	0.808	0.806	0.765

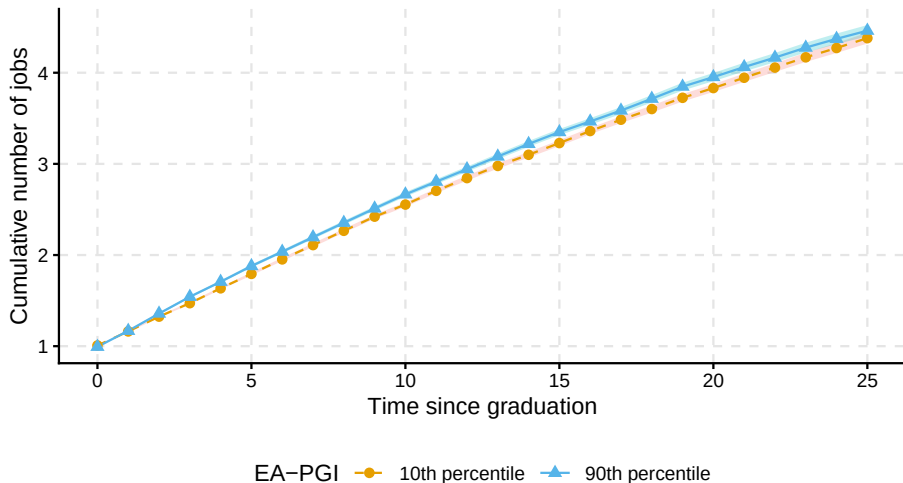
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

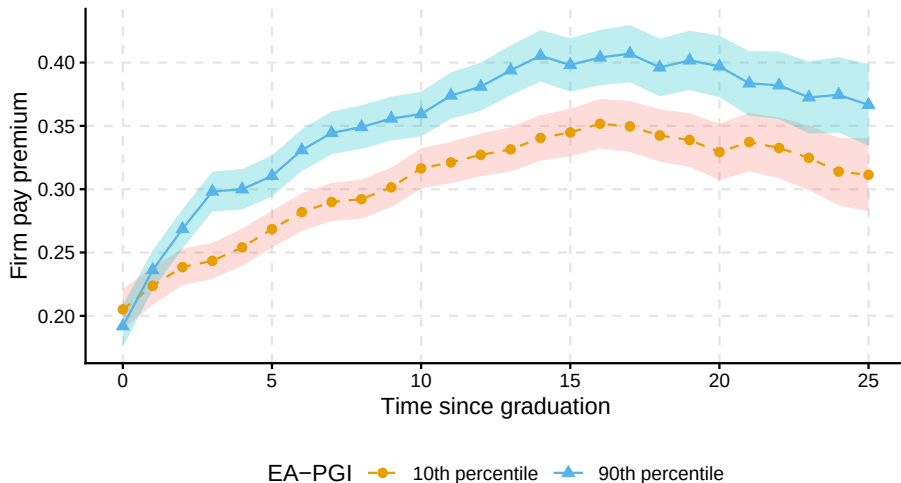


Mechanisms: Firm mobility

High EA-PGI individuals change jobs more rapidly



High EA-PGI individuals transition to higher-paying firms





Earnings growth decomposition (Hahn et al. 2021)

Accounting framework

$$\Delta \bar{y}_t = \underbrace{E_S \overline{s_t \Delta y_t}}_{\text{stayers}} + \underbrace{E_Q \overline{q_t \Delta y_t}}_{\text{employer-to-employer}} + \underbrace{E_N (\overline{n_t y_t} - \tilde{y}_t)}_{\text{entrance from non-empl}} - \underbrace{E_R (\overline{r_t y_{t-1}} - \tilde{y}_t)}_{\text{exit to non-empl}}$$

y_{it} log earnings of worker i at time t

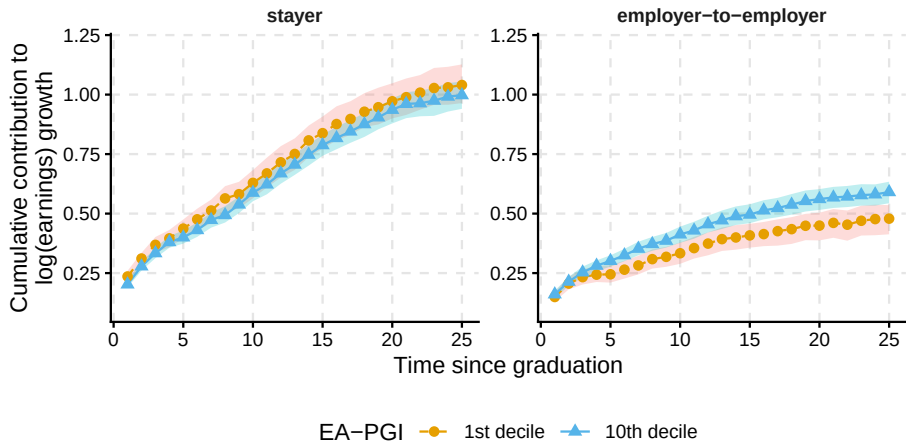
$s_{it} + q_{it} + n_{it} + r_{it} = 1, \forall i, t$

E_k employment share of worker type k

$\tilde{y}_t$ average income of stayers and employer-to-employer movers

Contribution of each mobility type to aggregate earnings growth

Income gap by EA-PGI attributable to job changes



► Non-employment mobility

Tertiary educated only; block bootstrapped CIs with 500 iterations



Mechanisms: Gelbach decomposition



Gelbach decomposition

Main: $y_{icmt} = \alpha_1 + \beta_{1t}PGI_i + \gamma_1X_{icmt} + \varepsilon_{icmt}$

Augmented: $y_{icmt} = \alpha_2 + \beta_{2t}PGI_i + \gamma_2X_{icmt} + \delta_2Z_{it} + \xi_{icmt}$

Auxiliary: $Z_{it} = \alpha_3 + \beta_{3t}PGI_i + \gamma_3X_{icmt} + u_{it}$

Mediator contributions $\beta_{1t} - \beta_{2t} = \sum \delta_2 \odot \beta_{3t}$

- Parents' PGI
- Education level and field
- Health index CCI
- Worker pay premium
- Firm pay premium, Occupation, and Firm pay premium $\times$ Occupation
- Industry and Firm tenure

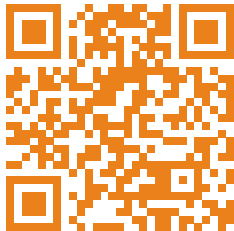
Worker and firm pay premiums account for most of the gap

	Time since graduation (years)				
	1–5	6–10	11–15	16–20	21–25
Main (EUR)	1 699 (558)	3 424 (629)	4 570 (730)	6 439 (997)	8 943 (1 447)
Unexplained	-53 (391)	-1 207 (620)	-481 (571)	-1 394 (830)	85 (1 040)
Mother's EA-PGI	-145 (203)	565 (288)	484 (297)	351 (458)	73 (481)
Father's EA-PGI	142 (502)	741 (488)	278 (279)	661 (526)	243 (546)
Education level	463 (65)	728 (88)	646 (85)	663 (117)	643 (133)
Education field	-393 (92)	-103 (77)	1 (54)	121 (109)	179 (149)
Health (CCI)	1 (9)	-3 (10)	0 (8)	1 (9)	5 (20)
Worker pay premium	1 357 (340)	2 114 (410)	2 838 (427)	5 081 (701)	5 883 (1 025)
Firm and occupation subtotal	374 (201)	527 (218)	865 (248)	872 (366)	1 646 (486)
Firm pay premium	35 (74)	98 (105)	187 (120)	282 (158)	180 (203)
Occupation (2-digit)	90 (116)	69 (178)	184 (141)	-220 (275)	431 (377)
Firm pay premium × occupation	249 (170)	360 (158)	493 (175)	810 (277)	1 036 (338)
Industry	-23 (78)	35 (91)	-38 (72)	75 (177)	183 (158)
Firm tenure	-23 (14)	26 (20)	-24 (18)	8 (37)	3 (40)
Observations	12 236	17 490	16 366	13 218	7 642



Conclusion

- Growing income gap by genetic predisposition over the lifecycle
- Indirect genetic associations and parental background highly relevant
 - Direct effect of own genes ↓ by 53%
 - Large part of EA-PGI channel explained by fathers (Del Boca et al. 2013)
- Potential mechanisms:
 - Sorting and differential returns to education
 - Large income gap attributed to transitions towards higher-quality employers
 - No sorting into first employer: uncertainty about match quality on both sides
 - Employer learning and job mobility become more important over time
- Weak association with health indices [▶ Graph](#)



Stay tuned for the updated draft!

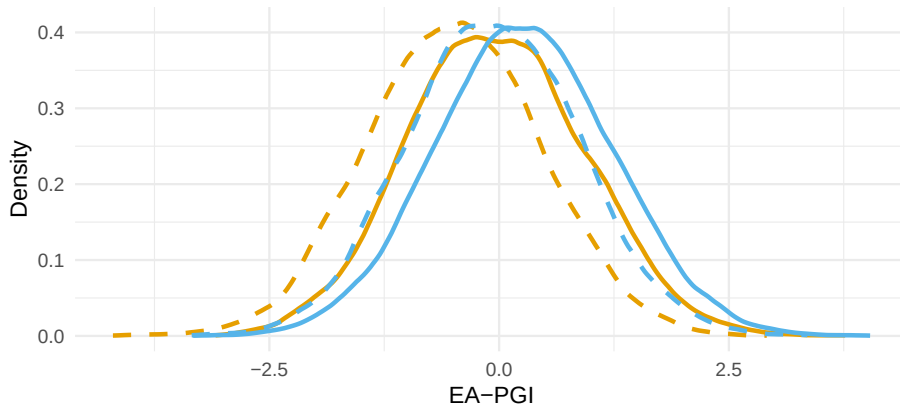
Nurfatima Jandarova nurfatima.jandarova@tuni.fi

Stefano Lombardi stefano.lombardi@vatt.fi



Appendix

EA-PGI distribution by highest education



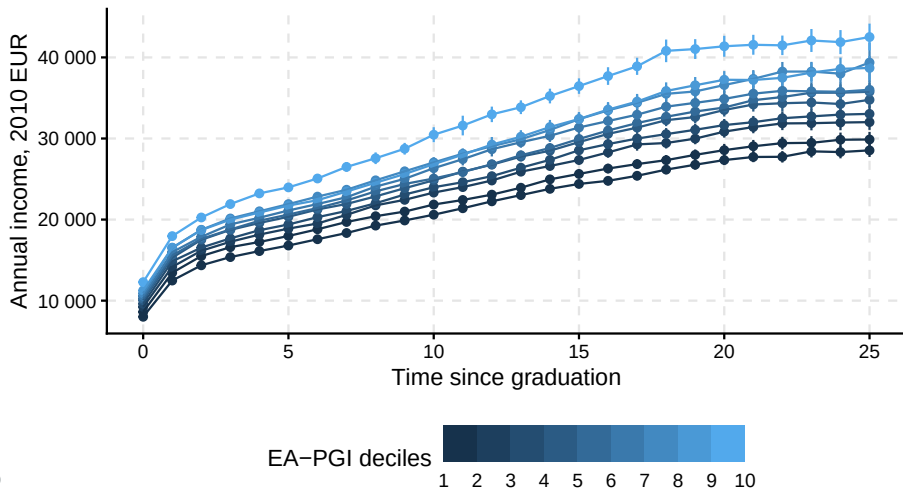
— Secondary academic — Tertiary academic
- - Secondary vocational - - Tertiary vocational



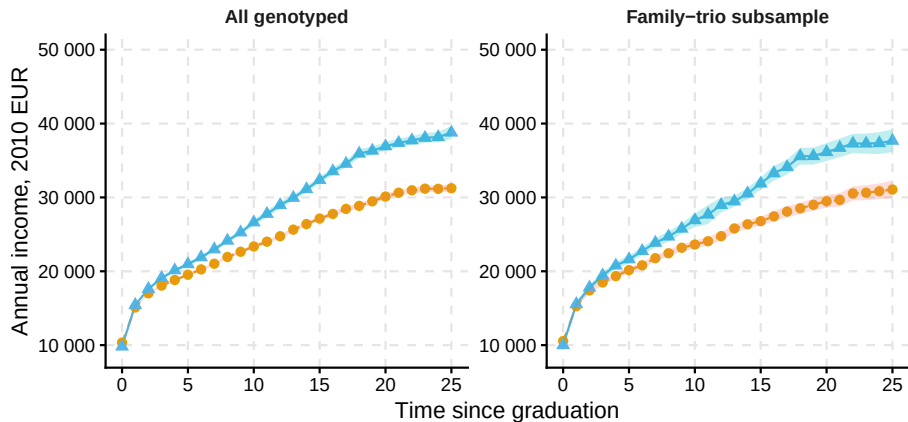
Balance table of genotyped graduates

	Population	Genotyped		Reweighted	
	Mean	Mean	p-val	Mean	p-val
Cohort: 1960-69	0.17	0.22	0.000	0.16	1.000
Cohort: 1970-79	0.34	0.36	0.000	0.34	1.000
Cohort: 1980-89	0.36	0.29	0.000	0.37	1.000
Cohort: 1990-99	0.13	0.11	0.000	0.13	1.000
Graduation age: 16-20	0.36	0.31	0.000	0.36	1.000
Graduation age: 21-25	0.39	0.43	0.000	0.39	1.000
Graduation age: 26-30	0.25	0.25	0.001	0.24	1.000
Education: secondary	0.44	0.37	0.000	0.44	1.000
Education: tertiary	0.56	0.63	0.000	0.56	1.000
Male	0.48	0.39	0.000	0.48	1.000
Married	0.10	0.13	0.000	0.11	0.000
Rural	0.24	0.25	0.712	0.25	1.000
Income at t=0	9 301	9 527	0.000	9 328	1.000

Average income trajectory by EA-PGI deciles



Baseline results in family trio subsample



Years of education in family analysis

	Baseline without parental EA-PGI		Controlling for parental EA-PGI	
	All family trios	Directly genotyped	All family trios	Directly genotyped
Own EA-PGI	0.553*** (0.016)	0.570*** (0.027)	0.413*** (0.026)	0.441*** (0.040)
Mother EA-PGI			0.128*** (0.021)	0.110*** (0.030)
Father EA-PGI			0.093*** (0.021)	0.095*** (0.030)
Constant	14.691*** (0.491)	13.741*** (1.058)	14.641*** (0.482)	13.717*** (1.028)
Obs.	12 871	4 586	12 871	4 586

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$



Weighted income gap

	Full sample		Family trio subsample			
	Baseline		Baseline		Controlling for parents' EA-PGI	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
10th percentile	309 659 (1 306)	291 728 (1 286)	295 550 (2 572)	295 123 (2 863)	303 725 (3 906)	304 546 (4 417)
50th percentile	329 893 (857)	308 756 (832)	313 108 (1 693)	313 741 (1 973)	313 190 (1 697)	313 886 (1 981)
90th percentile	350 418 (1 591)	325 930 (1 525)	330 870 (3 148)	332 602 (3 763)	322 764 (3 934)	323 347 (4 502)
Obs.	51 056	51 056	12 871	12 871	12 871	12 871

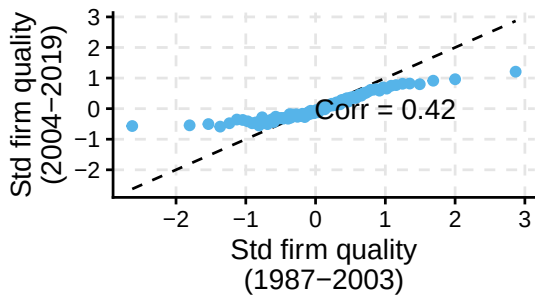
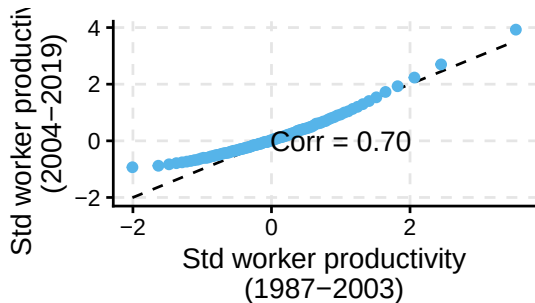
AKM summary statistics



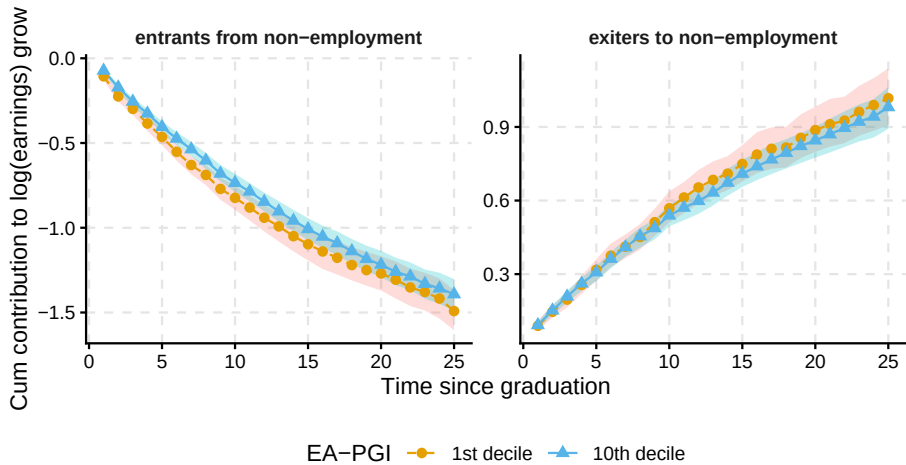
	1987-2003	2004-2019
Standard deviation of outcome	0.5003	0.4614
N estimation sample	16 586 748	15 060 995
N worker FE	1 881 715	1 842 564
N firm FE	126 605	50 430
<i>Panel A: Summary of parameter estimates</i>		
RMSE	0.1693	0.1669
Adjusted R ²	0.8846	0.8681
Worker FE	0.3547	0.4868
Firm FE	0.0458	0.0499
<i>Panel B: Share of outcome variance attributed to</i>		
Cov(worker FE, firm FE)	0.0269	0.0778
Xb and associated covariances	0.4712	0.2703
Residual	0.1014	0.1153



AKM fixed effects correlation

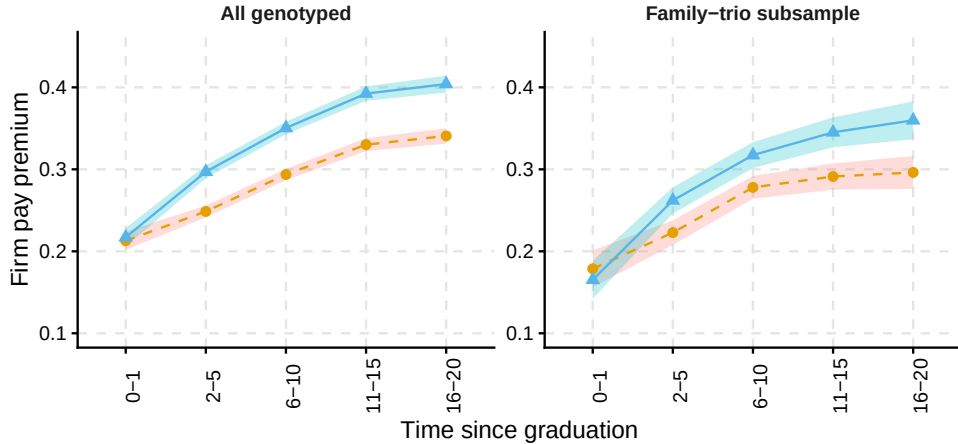


Contribution of non-employment mobility to earnings growth

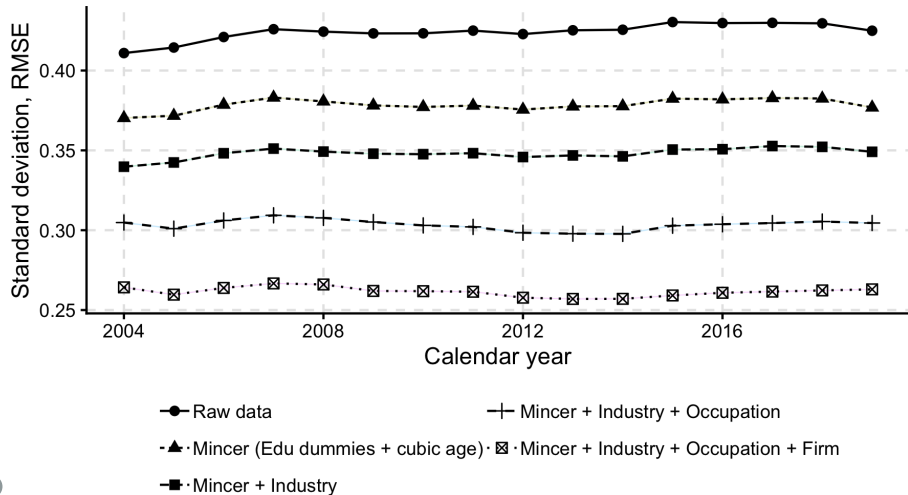




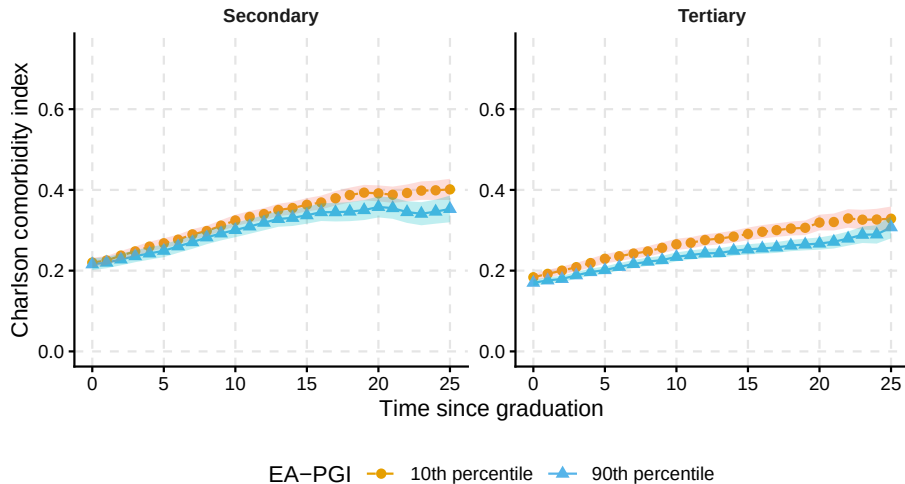
Baseline estimates of firm pay premium gap by EA-PGI



Income inequality, firms and occupations



EA-PGI weakly associated with health trajectories





References i

- Barth, Daniel, Nicholas W. Papageorge, and Kevin Thom. 2020. "Genetic Endowments and Wealth Inequality." *Journal of Political Economy* 128 (4): 1474–522. <https://doi.org/10.1086/705415>.
- Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36 (S1): S13–70. <https://doi.org/10.1086/694153>.
- Carvalho, Leandro S. 2025. "Genetics and Socioeconomic Status: Some Preliminary Evidence on Mechanisms." *Journal of Political Economy Microeconomics* 3 (3): 429–76. <https://doi.org/10.1086/732835>.



References ii

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47.

<https://www.aeaweb.org/articles?id=10.1257/aer.97.2.31>.

Del Boca, D., C. Flinn, and M. Wiswall. 2013. "Household Choices and Child Development." *The Review of Economic Studies* 81 (1): 137–85.

<https://doi.org/10.1093/restud/rdt026>.



References iii

- Ghirardi, Gaia, Carlos J. Gil-Hernández, Fabrizio Bernardi, Elsje van Bergen, and Perline Demange. 2024. "Interaction of Family SES with Children's Genetic Propensity for Cognitive and Noncognitive Skills: No Evidence of the Scarr-Rowe Hypothesis for Educational Outcomes." *Research in Social Stratification and Mobility* 92 (August): 100960. <https://doi.org/10.1016/j.rssm.2024.100960>.
- Hahn, Joyce K., Henry R. Hyatt, and Hubert P. Janicki. 2021. "Job Ladders and Growth in Earnings, Hours, and Wages." *European Economic Review* 133 (April): 103654. <https://doi.org/10.1016/j.euroecorev.2021.103654>.
- Kline, Patrick. 2024. *Firm Wage Effects*. Vol. 5. Elsevier. <https://doi.org/10.1016/bs.heslab.2024.11.005>.



References iv

- Okbay, Aysu, Yeda Wu, Nancy Wang, et al. 2022. "Polygenic Prediction of Educational Attainment Within and Between Families from Genome-Wide Association Analyses in 3 Million Individuals." *Nature Genetics* 54 (4): 437–49. <https://doi.org/10.1038/s41588-022-01016-z>.
- Rimfeld, Kaili, Eva Krapohl, Maciej Trzaskowski, et al. 2018. "Genetic Influence on Social Outcomes During and After the Soviet Era in Estonia." *Nature Human Behaviour* 2 (4): 269–75. <https://doi.org/10.1038/s41562-018-0332-5>.
- Rustichini, Aldo, William G. Iacono, James J. Lee, and Matt McGue. 2023. "Educational Attainment and Intergenerational Mobility: A Polygenic Score Analysis." *Journal of Political Economy* 131 (10): 2724–79. <https://doi.org/10.1086/724860>.

References v



Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019.
“Firming up Inequality*.” *The Quarterly Journal of Economics* 134 (1): 1–50.
<https://doi.org/10.1093/qje/qjy025>.