Inequality and Genes (and Family Background)

Markus Jäntti

Swedish Institute for Social Research
Stockholm University

1 februari 2018

Introduction

- intergenerational associations and the importance of family background in economic outcomes
- "genes":
  - genome-wide association studies ("GWAS")
  - population genetic models
- (economic) outcomes:
  - abilities (cognitive, socio-emotional ["non-cognitive"])
  - income (disposable family income, earnings, ...)
  - education
- inequality: the distribution of economic outcomes

Intergenerational economic associations

- suppose y_O and y_P are the "permanent income" of a pair of offspring and parent
- the intergenerational income elasticity is the measure for which most evidence is available:

$$y_O = \alpha + \beta y_P + \epsilon \tag{1}$$

- two interpretations for β:
  - the slope of the conditional expectation of offspring income, given parental income ("mechanical"):

$$\beta := \frac{\partial \mathrm{E}[y_O | y_P]}{\partial y_P} \tag{2}$$

  - the causal effect of a change in parental income on child income ("economic"):

$$\beta := \frac{\partial y_O^*}{\partial y_P} \tag{3}$$

    the y_O^* conveys that offspring income is at least in part the result of optimizing behavior on the part of parents

The causal interpretation

- the Becker och Tomes (1979, 1986) model of parental investment in child human capital inspirs much empirical work
- a simplified version is due to Solon (2004), with offspring income depending on parental income by

$$y_{i,O} = \mu^* + [(1-\gamma)\theta p] y_{i,P} + p e_{i,o}. \qquad (4)$$

  - p is the return on human capital
  - e is offspring human capital endowment
  - γ measures the progressivity in human capital
  - θ measures how effectively human capital investments turn into capital
  - λ captures the IG transmission of the ability (such as genetic transmission)

The causal interpretation

- in "steady state", the IGE is

$$\beta = \frac{(1-\gamma)\theta p + \lambda}{1 + (1-\gamma)\theta p \lambda} \tag{5}$$

- the intergenerational persistence *increases* in
  - the productivity of human capital investments θ
  - the income or earnings return to human capital p
  - the *heritability* of human capital endowments λ

  and *decreases* with
  - progressivity of public education spending γ
- the same factors drive cross-sectional inequality
- therefore IGE is also positively correlated with cross-section inequality [the "Great Gatsby curve" (Corak, 2013; Krueger, 2012)] Go to "Great Gatsby curve"

Cross-national results

- IGEs: 0.15-0.50 (acc. to Corak's (2013) version of the Great Gatsby Curve)
- IGCs: possibly less variation (Corak, Lindquist och Bhaskar Mazumder, 2013) but there is less comparable information about IGCs.
- Thus: R-squares (IGC^2) from 0.02-0.25.

Sibling correlation

- The prototypical model:

$$Y_{ij} = a_i + b_{ij}, \quad a \perp b \tag{6}$$

  - the "family effect" a shared by sibling in family i, variance σ_a^2
  - the "individual effect" b unique to individual j in family i (orthogonal to a), variance σ_b^2
- the population variance of the outcome Y is

$$\sigma^2 = \sigma_a^2 + \sigma_b^2, \tag{7}$$

- the share of variance attributable to family background (its "R^2") is

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2} \tag{8}$$

  which coincides with the Pearson correlation for sibling pairs

A sibling correlation captures more than an intergenerational correlation (IGC)

- an omnibus measure – captures both observed and unobserved family background (and neighborhood) factors
- yet it is a lower bound, because siblings don't share everything from the family background
- moreover,
  sibling correlation = IGC^2 + other shared factors that are uncorrelated with parental Y

Brother correlations in earnings and income

Country	Estimate	Source
Denmark	0.20	Schnitzlein (2013)
China	0.57	Eriksson och Zhang (2012)
Finland	0.26	Österbacka (2001)
Germany	0.43	Schnitzlein (2013)
Norway	0.14	Björklund, Eriksson m. fl. (2002)
Sweden	0.32	Björklund, Jäntti och Lindquist (2009)
USA	0.49	Bhashkar Mazumder (2008)

Sibling correlations in years of schooling

Country	Sibling type	Estimate	Source
Germany	Brothers	.66	Schnitzlein (2013)
Germany	Sisters	.55	Schnitzlein (2013)
Norway	Mixed sexes	.41	Björklund och Salvanes (2011
Sweden	Brothers	.43	Björklund och Jäntti (2012)
Sweden	Sisters	.40	Björklund och Jäntti (2012)
USA	Mixed sexes	.60	Bhashkar Mazumder (2008)

These quite high numbers are only lower bounds. What is missing?

1. differential treatment by parents. Will not be captured if it creates differences, but is part of family background.
2. full siblings have only about half of (initial) genes in common. But each individual has 100% of her (initial) genes from the parents.
3. not all environmental experience and “shocks” are shared, only some. Thus some environmental stuff is missing.

Sibling correlations vs. intergenerational correlations, Swedish estimates

- recall that:
  sibling correlation = IGC^2 + other shared factors that are uncorrelated with parental Y

	ρ	$\text{IGC}^2 = R^2$	Other factors
Brothers:			
Earnings	.24	.02	.22
Schooling	.46	.15	.31
Sisters:			
Schooling	.40	.11	.29

Genetics and inequality

See Beauchamp m. fl. (2011) och Manski (2011)

Two types of approaches:

- modern: genome-wide association studies and inequality
- traditional: population-genetic modelling

Genome-wide association studies and inequality

See e.g. Beauchamp m. fl. (2011) och Chabris m. fl. (2012)

- linking genetic markers/single nucleotide polymorphisms (SNPs) to specific (economic) traits
- a quickly moving and expanding field of study ...
- ... that yields both many insights but also many disappointments ...
- ... but one which as of yet has yielded very few insights into the genetic basis of economic inequality

GWAS is providing information from the research frontier, but now mostly providing insights into the associations with the *levels* of economic traits rather than with the *inequality* of economic outcomes.

Cautionary note

From "Most Reported Genetic Associations With General Intelligence Are Probably False Positives", (Chabris m. fl., 2012)

General intelligence (g) and virtually all other behavioral traits are heritable. Associations between g and specific single-nucleotide polymorphisms (SNPs) in several candidate genes involved in brain function have been reported. We sought to replicate published associations between g and 12 specific genetic variants [. . .] using data sets from three independent, well-characterized longitudinal studies with samples of 5,571, 1,759, and 2,441 individuals. Of 32 independent tests across all three data sets, only 1 was nominally significant. By contrast, power analyses showed that we should have expected 10 to 15 significant associations, given reasonable assumptions for genotype effect sizes. [. . .] We conclude that the molecular genetics of psychology and social science requires approaches that go beyond the examination of candidate genes.

Population genetic models [PGM] and inequality

- started long before the role of molecular genetics was well understood
- relies (often) on studies of twins (MZ/DZ, reared together/apart) but can rely on general kinship
- an important aim has been to estimate the extent to which variation in some trait (IQ; personality measures; education; income) is genetic (heritability)
- relies on

$$\text{outcome} = \text{genetic factors} + \text{environmental factors} \qquad (9)$$

  or

$$Y = G + E$$

- "environmental factors" E are further separated into "shared" ones (S; such as the behaviour of parents toward their children) and non-shared ones (U)

Illustrative example: PGM for earnings in Sweden

Björklund, Jäntti och Solon (2005)

- strategy: estimate highly restricted, unrealistically simple model and extend it gradually (ad hoc)
- simple model of earnings determination:

$$Y = gG + sS + uU \tag{10}$$

- Y is permanent (=long-run) earnings. *Normalize* the variance of Y to unity.
- G, S and U are additive gene effect, shared and non-shared environment that are unobserved, latent variables. *Normalized* to have unit variance and zero mean.
- g, s and u are "factor loadings", parameters to be estimated. Interest in g^2 ("heritability") and s^2 in particular.
- by assumption, the population variance in Y is

$$\text{Var}(Y) \equiv 1 = g^2 + s^2 + u^2. \tag{11}$$

- the parameters g^2 and s^2 can be identified by correlations in Y between relatives.
- let Y and Y' be two related persons:

$$\begin{aligned}\mathrm{Cov}(Y, Y') = g^2\mathrm{Cov}(G, G') + s^2\mathrm{Cov}(S, S') + u^2\mathrm{Cov}(S, S') + \\ 2gb\mathrm{Cov}(G, S') + 2gu\mathrm{Cov}(G, U') + 2su\mathrm{Cov}(S, U').\end{aligned} \tag{12}$$

- in order to estimate these parameters, we must place a number of restrictions on the covariances of the latent variable.
- assume non-shared environment U un-correlated with everything

$$\mathrm{Cov}(G, U') = \mathrm{Cov}(S, U') = \mathrm{Cov}(U, U') = 0 \tag{13}$$

- if mating is random, there are no dominant gene effects nor non-additive gene effects, $\mathrm{Cov}(G, G')$ is 1, .5 and .25 for identical twins, fraternal twins as well as full siblings and half siblings
- for siblings reared together, $\mathrm{Cov}(S, S')$ is 1, 0 otherwise
- focus here on brother only (the paper reports results for both brothers and sisters)

”Design matrix” for estimating variance components from sibling correlations

Sibling type	Rearing	$\mathrm{Cov}(G, G')$	$\mathrm{Cov}(S, S')$
	Model 1		
MZ twins	Together	1	1
DZ twins	Together	0.5	1
MZ twins	Apart	1	0
DZ twins	Apart	0.5	0
Full sibs	Together	0.5	1
Half sibs	Together	0.25	1
Full sibs	Apart	0.5	0
Half sibs	Apart	0.25	0
Adopted	Together	0	1

Three variations to the simple model ($1 \to 2\{A,B,C\}$)

A Gene-env correlation

- replace the assumption that $\mathrm{Cov}(G, S') = 0$ with parameters to be estimated
- one for biological siblings reared together, one for those reared apart, one for adoptive siblings

B Gene-gene correlation

- suppose that some of the restrictive assumptions that generate $\mathrm{Cov}(G, G')$ of 1, .5, .25 are violated
- allow correlations be different for identical twins, fraternal twins, half siblings and adoptive siblings

C Shared environment

- suppose “rearing”, or the shared environment is not on, off
- normalize $\mathrm{Cov}(S, S')$ for MZ twins reared together, one for fraternal twins together, one for other siblings reared together and one for siblings reared apart

Consequences of changes to assumptions for MZ/together genetic/environment components

	"Raw" correlation	Genetic component	Environmental component
Model 1	.363	.281	.038
2A Vary G, S	.363	.250-.314	.020-.084
2B Vary G, G'	.363	.320	.037
2C More env sim for MZ	.363	.199	.164

Additional remarks

- *in utero* shocks are now known to be important ...
- ...but the PGM assigns all pre-birth factors to genes
- PGM models tend to normalize the variances and work only with correlations ...
- ...which, by construction, abstracts from distributional dynamics
- however, population genetic analysis of family associations provides useful insights and structure to understanding family associations in outcomes

See, e.g., Kamin och Goldberger (2002), Feldman, Otto och Christiansen (1999).

Policy implications of PGM

See Goldberger (1979) och Manski (2011)

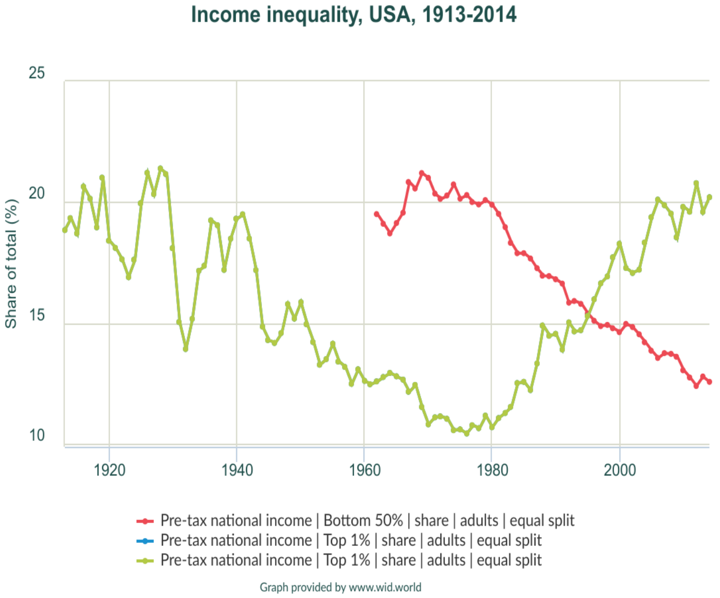
- near-sightedness (aka. myopia)
- it seems reasonable to suppose that myopia is highly heritable
- does it follow that there is no appropriate policy response to it?
- no; distributing eyeglasses to the myopic is an effective way to alleviate nearsightedness
- the fact of high heritablility tells us little or nothing of how amenable a disadvantage is to *interventions*
- for that, we need estimates of the *causal effect* of interventions

”The conclusion is that the heredity-IQ controversy has been a ’tale full of sound and fury, signifying nothing’. To suppose that one can establish effects of an intervention process when it does not occur in the data is plainly ludicrous.” [Kempthorne, 1978, cited by Manski (2011))]

Concluding remarks

- population and molecular genetics will be continued to be explored by social scientists ...
- ...and the latter are likely to increasingly provide insights into the scope for policy interventions to be effective
- the *dynamics* of economic inequality, the extent to which there is equality of opportunity, continue to be of great interest
- many interesting issues to be studies *apart* from the role of genetics in these processes
  - e.g., what are the things that families do that is *not* captured in direct IG transmission?
  - can and do policy interventions alter the strength of family associations?

Trends in income inequality



Policy interventions and family associations

- public expenditures in US affects IGE (Mayer och Lopoo, 2008)
- comprehensive school reform in Sweden, Finland had sizeable impact on IGE

Comprehensive school reform in Finland

- Comprehensive school thus:
  - moved tracking from age 11 to age 16
  - increased length of compulsory schooling by one year
  - led to integration of students in same schools between ages 11-16
  - made all follow same curriculum between ages 11-16 (although some variation initially)
- The reform was implemented in 5 stages between 1972 and 1977, affecting cohorts born 1961-1965, starting in the north and ending in the capital area.
- We use the stage-wise implementation to estimate the effect of the reform by comparing the correlation among pairs of brothers who either were or were not affected by comprehensive school

The impact of comprehensive school reform in Finland

Father's earnings	0.277	0.297	0.298
	(0.014)	(0.011)	(0.010)
Father's earnings$\times$Reform		-0.055	-0.069
		(0.009)	(0.022)
Reform		-0.065	-0.019
		(0.012)	(0.021)

Source: Pekkarinen, Uusitalo och Kerr (2009)

Inequality is on the increase

Average annual growth across the income distribution ca 1985-2008 (before the Great Recession) [Source: OECD (2011)]

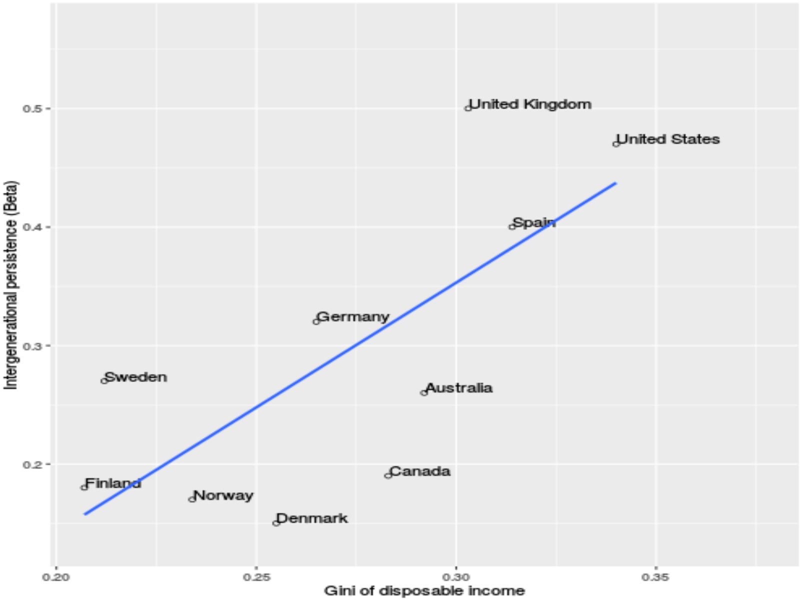
	%-change		
	Overall	Bottom 10%	Top 10 %
Australia	3.6	3.0	4.5
Austria	1.3	0.6	1.1
Canada	1.1	0.9	1.6
Denmark	1.0	0.7	1.5
Finland	1.7	1.2	2.5
France	1.2	1.6	1.3
Germany	0.9	0.1	1.6
Italy	0.8	0.2	1.1
Mexico	1.4	0.8	1.7
Netherlands	1.4	0.5	1.6
Norway	2.3	1.4	2.7
Portugal	2.0	3.6	1.1
Spain	3.1	3.9	2.5
Sweden	1.8	0.4	2.4
United Kingdom	2.1	0.9	2.5
United States	1.3	0.5	1.9
OECD27	1.7	1.3	1.9

What if the Great Gatsby curve persists while inequality increases?

- the “Great Gatsby” curve plots the intergenerational persistence of income against income inequality in (roughly) the parental generation
- income inequality has increased
- what can be expected of persistence?
- **caveat:** this is highly speculative and is intended as food for thought

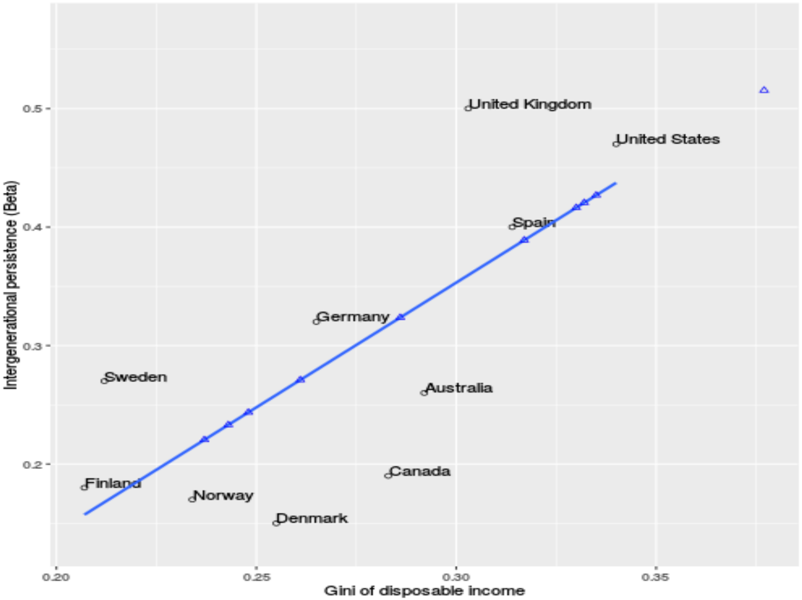
The expected evolution of income persistence

GG curve for subset of countries in Corak (2013) also in Luxembourg Income Study



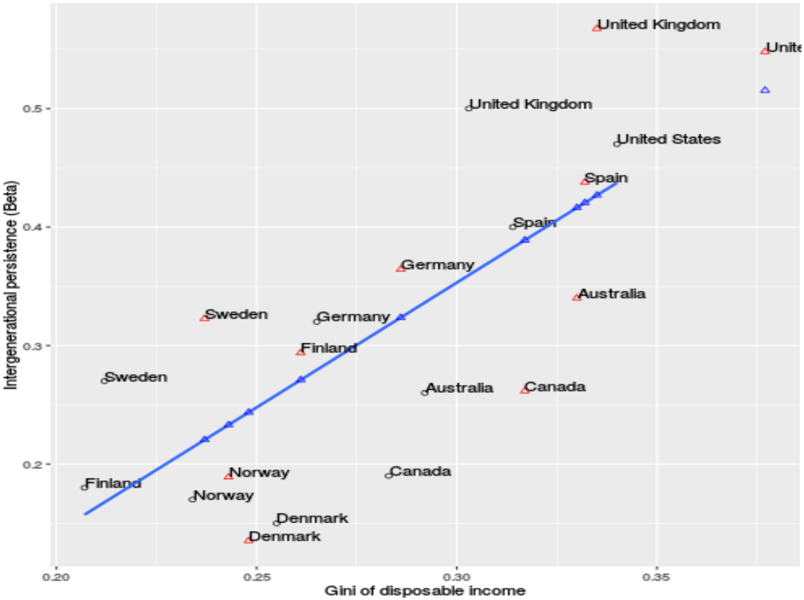
The expected evolution of income persistence

GG curve for subset of countries in Corak (2013) also in Luxembourg Income Study



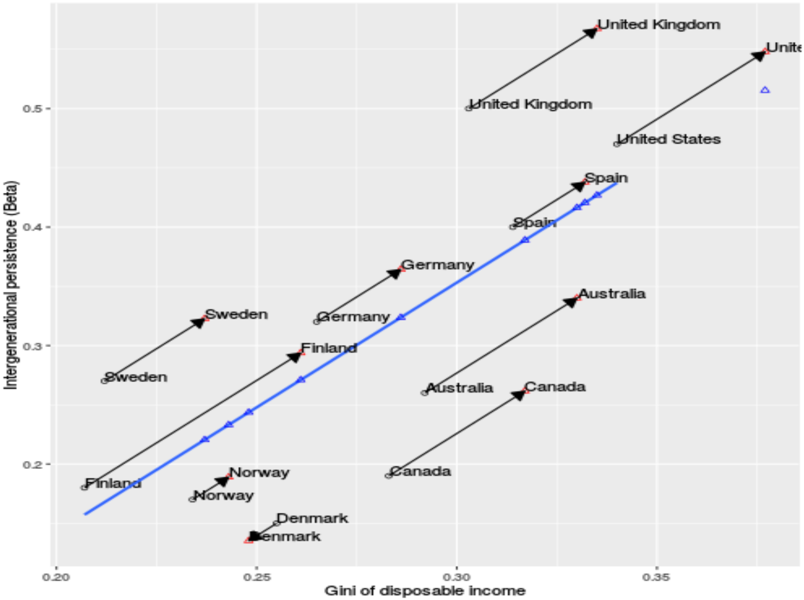
The expected evolution of income persistence

GG curve for subset of countries in Corak (2013) also in Luxembourg Income Study



The expected evolution of income persistence

GG curve for subset of countries in Corak (2013) also in Luxembourg Income Study



The Great Gatsby curve

the relationship between intergenerational earnings persistence and cross-sectional income inequality; Source: Corak (2013, Figure 1). ▸ Go back to Causal intrepretation