Economics & Genetics Lecture 1

DR. NIELS RIETVELD, NRIETVELD@ESE.EUR.NL
OFFICE MANDEVILLE T18-29

Today's agenda

Main questions:
- Why should economists be excited about genetics?
- How to estimate heritability using twins?
- Is the estimation of heritability policy-relevant?

Literature:
- Nicolaou et al. (2008): "Is the tendency to engage in entrepreneurship genetic?"
- Manski (2011): "Genes, eyeglasses, and social policy"

Economics & Genetics Lecture 1 – Part 1

MAIN QUESTIONS 1 & 2

Why should economists be excited about genetics?

(Micro-)economists try to explain (economic) decisions people make, and analyze the consequences of these choices

- For example, the impact of schooling ("investments in human capital") on later life outcomes

Social sciences (and economics as part of it) traditionally focus on "environmental" factors to explain individual (socio-economic) differences

- "The liberal orthodoxy back then was – at the risk of some exaggeration, but not much – that *anyone could learn anything*, and that the only fair system was to put students with completely different levels of competence and knowledge into the same classroom." (Quote about the 1970s)

Humans beings are not "blank slates": Many heritability studies have shown that *genes* do impact schooling performance

→ "The nature vs. nurture debate is over" (Turkheimer, 2000)

→ When economists ignore genes, they will never be able to provide complete explanations of choice behavior

The three laws of behavioral genetics

Turkheimer (2000): "The nature vs. nurture debate is over":

1. **All human behavioral traits are heritable** *(a trait is a specific characteristic of an organism)*
2. **The effect of being raised in the same family is smaller than the effect of genes**
3. **A substantial portion of the variation in complex human behavioral traits is not accounted for by the effects of genes or families**

Three explanatory factors: Genes – Family environment – Unique environment

Heritability of different traits

Think about the following two questions:

1) How would you define heritability?

2) Order the following traits in the order of decreasing heritability:

- BMI
- Happiness
- Huntington disease
- Educational attainment
- Height

Meaning of heritability

Definition of heritability:

- *The* <u>*proportion*</u> *of observed differences in a trait among individuals from a certain population that is due to genetic differences among these individuals*

Some caveats:

- *Differences:* A trait that does not vary in a population may be *inherited* without being *heritable* (e.g., having two legs)
- *Population* parameter: Not the same as saying the "fraction X" of a trait of *a person* is caused by genetic factors
- Population specific: Not a natural constant across populations
- Uninformative about *genetic* architecture: Few genes with large effects? Many genes with small effects?

Heritability of different traits

Variance decomposition: Heritability is higher when environmental influences are smaller

Heritability may be *induced* or *reduced* by specific features of the environment

- For example, a society ("dictator") that sends all girls to university and all men to the army
  - Heritability induced by environment
- For example, heritability of height is relatively low in poor countries
  - Scarcity of food reduces heritability

Huntington disease = 100% ("monogenetic disease")
Height = 90%
BMI = 70%
Educational attainment = 40%
Happiness = 30%

Heritability of different traits

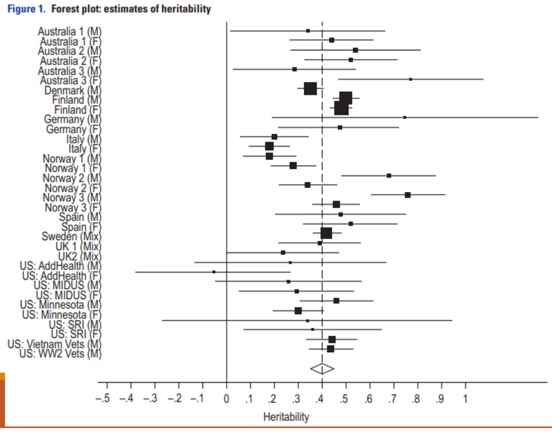
Polderman et al. (2015): "Meta-analysis of the heritability of human traits based on fifty years of twin studies"

- Meta-analysis based on 2,748 publications, and 14,558,903 partly dependent twin pairs
- 17,804 traits studied (e.g., antropometric, psychiatric, cognitive, behavioral…);
- Virtually all traits that were ever investigated (with twin studies) showed an influence of genes

Average heritability across all traits: 49%!

Branigan et al. (2013): "Variation in the heritability of educational attainment: An international meta-analysis"

Figure 1. Forest plot: estimates of heritability



Estimating heritability – twin studies

The challenge: Separate gene effects from family environment effects

- Possible with adoption studies, but samples usually small and non-representative

Classic solution: Twin study

- Compare monozygotic (MZ) and dizygotic (DZ) twins reared together
- The most important assumptions:
  1. Genetic correlation is 1 for MZ twins, and 0.5 for DZ twins *(Lecture 3)*
  2. "Equal environment assumption": Family environment is not more similar for MZ twins than for DZ twins (usually OK)
  3. No assortative mating of parents on the trait in question *(otherwise assumption 1 not correct, see Tutorial 1)*

Zygosity

DNA zygosity test

vs.

observed similarity test

Peas-in-a-Pod questionnaire: twin adults

The following questions ask how alike you and your twin were in your childhood. Please choose the most fitting answer for each question.

Were you and your twin "as alike as two peas in a pod"?

1. As alike as two peas in a pod
2. Usual sibling similarity
3. Quite different

Were you and your twin mixed up as children?

1. Yes, very often
2. Now and then
3. Never

By whom were you mixed up?

1. Parents
2. Teachers
3. Others
4. Nobody

The answers for each question are numbered. Add the answer number for each of your chosen answers to get a score from 3 to 10. This questionnaire is most accurate if the scores from you and your twin's results are added together to get a score from 6 to 20.

A combined score of 6 to 13 indicates you and your twin would be monozygotic.

A combined score of 14 to 20 indicates you and your twin would be dizygotic.

Using the combined score from you and your twin's results has a 93% chance of indicating your true zygosity. If you would like to confirm your zygosity we recommend a DNA zygosity test.

Estimating heritability – twin studies

Estimate heritability using MZ and DZ twins

- Heritability is "the proportion of observed differences in a trait among individuals from a certain population that is due to genetic differences among these individuals"
- Focus on <u>differences</u> (not levels, as we for example usually do in regular regressions)
- *Decomposition* of variances (<u>independent</u> variance components):

$$Var(y) = Var(a) + Var(c) + Var(e)$$

- *y* – trait (e.g., educational attainment), *a* – <u>**a**dditive</u> genetic effects, c – **c**ommon (family) environment, *e* – unique **e**nvironment

(Narrow-sense: *Only additive genetic effects*) heritability $h^2 = Var(a) / Var(y)$

Estimating heritability – twin studies

With *Falconer's formula*, we can estimate heritability by exploiting the twin structure of the data

- **MZ** twins are similar because of their genes (they are 100% similar) and their family environment
- **DZ** twins are similar because of their genes (but they are only 50% similar) and their family environment

Trait correlation between MZ twins is $r_{MZ} = h^2 + c^2$

Trait correlation between DZ twins is $r_{DZ} = 0.5 \times h^2 + c^2$

The heritability of educational attainment

- **Trait** correlation between MZ twins is $r_{MZ} = h^2 + c^2$
- **Trait** correlation between DZ twins is $r_{DZ} = 0.5 \times h^2 + c^2$

Exercise: Try to re-arrange terms in such a way that you get an expression for the h^2

- *I.e.,* $h^2 = \ldots$

Branigan et al. (2013): $r_{MZ} = 0.70$, $r_{DZ} = 0.53$ (Australia)

- Estimate heritability using Falconer's formula

Estimating heritability – twin studies

Trait correlation between MZ twins is $r_{MZ} = h^2 + c^2$

Trait correlation between DZ twins is $r_{DZ} = 0.5 \times h^2 + c^2$

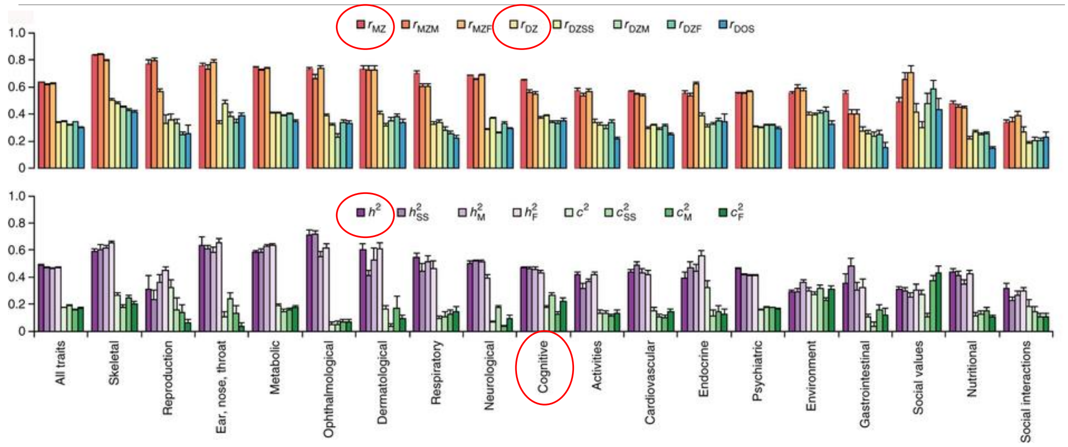
Subtract second equation from first equation:
- $r_{MZ} - r_{DZ} = h^2 + c^2 - 0.5 \times h^2 - c^2 = 0.5 \times h^2$
- $h^2 = 2 \times (r_{MZ} - r_{DZ})$

- $h^2 = 2 \times (0.70 – 0.53) = 0.34$

Common (shared) environmental effects:
- $c^2 = r_{MZ} - h^2 = r_{MZ} - 2 \times (r_{MZ} – r_{DZ})$
- $c^2 = -r_{MZ} + 2 \times r_{DZ}$

Unique (non-shared) environmental effects:
- $a^2 + c^2 + e^2 = 1$
- $2 \times (r_{MZ} - r_{DZ}) - r_{MZ} + 2 \times r_{DZ} + e^2 = 1$
- $e^2 = 1 - r_{MZ}$

Estimating heritability – twin studies



Source: Polderman et al. (2015)

Thought experiment

What if there is *assortative mating*?

- Parents select each other based on observable similarity
- Would this result in an overestimation or underestimation of heritability?

- Answer: Dizygotic twin pairs will be more similar than expected. This is problematic for the classical twin model, because…
  - …we assumed $r_{DZ} = 0.5 \times h^2 + c^2$ which implied $h^2 = 2 \times (r_{MZ} - r_{DZ})$
  - If genetic correlation is larger than 0.5, then we overestimate h^2 *(and underestimate c^2)*!

ACE modelling

Structural equation modelling (SEM), in the spirit of Falconer's formula

- You'll practice with it in Tutorial 1

"ACE" model: Quite often, one uses short-hand notation:

Var(a) / Var(y) = a^2 or h^2 or *A*

Var(c) / Var(y) = c^2 or *C*

Var(e) / Var(y) = e^2 or *E*

SEM model comes with flexibility:

- Works with trait correlations *or* individual-level data
- You can adjust for assortative mating
- Can give confidence intervals for your estimates
- Allows you to compare models (ACE to AE/CE) based on model fit (likelihood of the model)

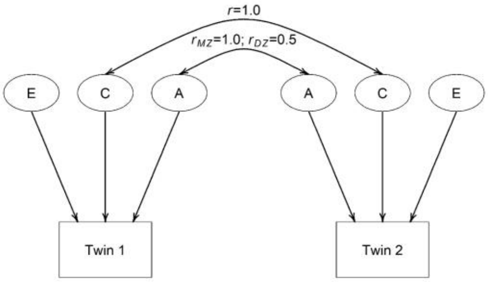
ACE modelling

Full ACE model, versus AE submodel and CE submodel

Occam's razor (idea attributed to English Franciscan friar William of Ockham (c. 1287–1347), a scholastic philosopher and theologian): *"Entities should not be multiplied without necessity"*

- "One should select the solution with the fewest assumptions"

A bit counterintuitive: If the fit of the data to the model (assessed by the change in likelihood) does **not** change significantly ($p > 0.05$), then the more parsimonious model is preferred



Is the tendency to engage in entrepreneurship genetic? (Nicolaou et al., 2008)

(Important motivation for start of my PhD project in 2010):

- If entrepreneurship is heritable, we should be able to find the underlying genes
- And if we know the genes, we can investigate whether it has implications for your health (stress) if your genes do not fit your occupation

Are entrepreneurs born or made?

- Entrepreneurship runs in families, e.g., family businesses
- Entrepreneurship courses do not really seem to have impact
- Entrepreneurship needs the right environment, e.g., Soviet Russia vs. US

Nicolaou et al. (2008) results

Data from TwinsUK (mostly females)

Table 4 Genetic Modeling Analysis to Predict Currently Self-Employed

Model	A (95% CI)	C (95% CI)	E (95% CI)	χ^2	df	*p*-value	AIC	RMSEA
ACE	0.48 (0.17 to 0.63)	0 (0 to 0.23)	0.52 (0.37 to 0.69)					
CE	—	0.34 (0.20 to 0.47)	0.66 (0.53 to 0.80)	11.74	4	0.02	3.74	0.05
AE	0.48 (0.31 to 0.63)	—	0.52 (0.37 to 0.69)	4.53	4	0.34	−3.47	0.01

Note. A, additive genetic; C, common environment; E, unique environment.

Change in fit of ACE model compared to CE model *significant* (*p* = 0.02)
- Hence, we are not willing to adopt the CE model

Change in fit of ACE model compared to ACE model *not significant* (*p* = 0.34)
- "Same fit", more parsimonious model: AE model is preferred over ACE model

Nicolaou et al. (2008) results

Heritability surprisingly similar for all empirical measures of entrepreneurship

Table 7 Summary of the Results of the Genetic Modeling Using Different Measures of Entrepreneurship

Dependent variables	A (95% CI)	C (95% CI)	E (95% CI)
Years self-employed	0.39 (0.24 to 0.44)	—	0.61 (0.56 to 0.66)
Owner operator	0.37 (0.25 to 0.49)	—	0.63 (0.51 to 0.75)
Number of companies owned and operated	0.37 (0.32 to 0.42)	—	0.63 (0.58 to 0.68)
Having started a business	0.41 (0.31 to 0.51)	—	0.59 (0.49 to 0.69)
Number of businesses started	0.42 (0.37 to 0.47)	—	0.58 (0.53 to 0.63)
Engaged in start-up effort	0.41 (0.31 to 0.50)	—	0.59 (0.50 to 0.69)
Number of start-up efforts	0.42 (0.37 to 0.47)	—	0.58 (0.53 to 0.63)
Attitudes toward entrepreneurship as a career	—	0.18 (0.13 to 0.23)	0.82 (0.77 to 0.87)

Notes. A, additive genetic; C, common environment; E, unique environment. The results for the best-fitting model for each variable are shown.

Exemplary results

Heritability studies show that entrepreneurs are born *and* made
- Surprisingly, common environment does not seem to be important
- Still, right unique environment matters quite a deal

Exemplary results for the three laws of behavioral genetics (Turkheimer):
1. All human behavioral traits are heritable
2. The effect of being raised in the same family is smaller than the effect of genes
3. A substantial portion of the variation in complex human behavioral traits is not accounted for by the effects of genes or families

Economics & Genetics Lecture 1 – Part 2

MAIN QUESTION 3

Is the estimation of h^2 policy relevant?

The classical twin study *decomposes* trait variance:

$$Var(y) = Var(a) + Var(c) + Var(e)$$

y – trait (e.g., educational attainment), *a* – **a**dditive genetic effects, c – **c**ommon (family) environment, *e* – unique **e**nvironment

(Narrow-sense: *Only additive genetic effects*) heritability $h^2 = Var(a) / Var(y)$

Importantly:

- We do not measure the environment: It's *everything* we cannot attribute to genes

Thought example

Inspired by Jencks (1980), "Heredity, environment, and public policy reconsidered"

- Imagine two genetically different strains (A and B) of corn in the same field
- The yield in this hypothetical field depends entirely on the number of grasshoppers attacking the crop
- Strain A yields more than Strain B solely because Strain A is less attractive to grasshoppers

Question: Are genes or the "environment" (the grasshoppers) determining variance in yield?

Exemplary illustration for the debate between "hereditarians" and "environmentalists"

- It matters which source you consider as most fundamental: "genes first" vs. "environment first" explanations
- *(In Lecture 5 we discuss this issue more in depth when we discuss gene-environment interactions)*

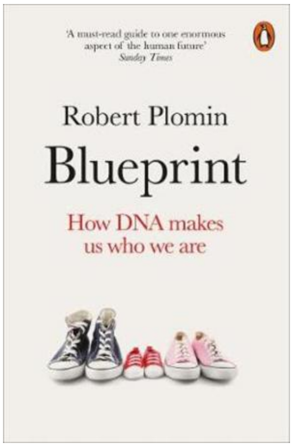
A heriditarian perspective on parenting

"Practical implications": Have heritability studies shown that "parenting is overrated"?

Watch a few minutes from a recent public debate about "Is parenting overrated" on October 29, 2019

5.57 – 8.01: https://www.youtube.com/watch?v=0hqjkBAd20M

Parenting important, but "…parents are responding to genetic differences in their children, rather than creating these differences"



"Is parenting overrated?"

Twin studies and the heriditarian *stance*

Twin studies take a "heriditarian perspective": "genes first"

- The A component in twin studies relates to everything that can *eventually* be traced back to genetics (Plomin: also parenting if it is a reaction to A)
  - Genetic component may capture systematic environmental channels ("grasshoppers") through which these genes operate

Reactive "gene-environment correlation": The environment responds differently to individuals with different genes (e.g., giving books to kids who enjoy reading, different treatment males/females)

Active "gene-environment correlation": Individuals with different genes seek out or create different environments (e.g., those with light skin avoid sunny environments, MZ twins reared apart often very similar)

Why still care about h^2?

Significant h^2's indicate that blank slate theories *(environment determines everything)* don't suffice

- Nevertheless, a "genes-first" conclusion is also incorrect as genetic effects may run through environments

It gives an upper bound on predictive power of genes in a sample

- $h^2 > 0$ necessary condition for gene discovery (Lecture 4)

However, is it relevant for policy to estimate h^2?

- This question relates to an unsettled debate from the 1970's of the previous century

The policy relevance of h^2

In particular the economist Paul Taubman showed a couple of times that the h^2 of income is 20-40%

- Hans Eysenck (psychologist): The British Commission for the Distribution of Income might as well "pack up" if genes determine so much of the income differences

The economist Arthur Goldberger disagreed completely

- Even if heritability = 100%, this does not imply irrelevance of the environment
- *It is also not true that* "…if it were shown that a large proportion of the variance in eyesight were due to genetic causes, then the Royal Commission on the Distribution of Eyeglasses might as well pack up." (Goldberger 1979, p. 337) → The provision of eye glasses may solve this genetically influenced eyesight problem (discussed by Manski, 2011)

- So, "Heritability analysis is just not a guide for policy, not a short-cut around the detailed cost-benefit analysis required for each specific policy proposal. On this assessment, heritability estimates serve no worthwhile purpose" (Goldberger, 1979, p. 346)

The policy relevance of h^2

Paul Taubman replied to Goldberger's study titled "Heritability" with a study titled "On heritability"

"Heritability studies are informative for policy analysis by providing *indicators for the presence of inequality of opportunity*"

  - Inequality of opportunity: Not all members of a population are eligible to compete on equal terms for same social position (e.g., access to specific schools, jobs, etc.)

Taubman reasoned that *family related factors* such as wealth, networks, etc. contribute to inequality of opportunity

- These factors all subsumed under c^2 not h^2 ($=a^2$) in a twin study
- But Taubman purposely decided to estimate heritability as ($a^2 + c^2$) rather than a^2 because he considered the family factors of importance

Revisiting the unsettled debate

Despite Taubman's reply, economists lost interest in h^2 studies…
- See Manski (2011), "Genes, eyeglasses, and social policy", *Journal of Economic Perspectives*, 25, 83-94.
- For Dutch description of this debate see my article "Geno-economie" in *Economisch Statistische Berichten*, 101, 180-183.

September 2020: Grant from European Research Council to study "Genes, policy, and social inequality"
- Estimation of h^2 still not **directly** relevant, but **indirectly**: Focus on c^2 before and after a policy change (Longitudinal heritability analysis)
- For example, has c^2 increased since student grants were turned into student loans?
- For the empirical analyses, I will merge registry data from Statistics Netherlands with zygosity information from the Netherlands Twin Registry (~122,500 twin pairs, and their nuclear and extended family members)

Moreover: Genoeconomics!

Manski (2011)

Abstract

Someone reading empirical research relating human genetics to personal outcomes must be careful to distinguish two types of work: An old literature on heritability attempts to decompose cross-sectional variation in observed outcomes into unobservable genetic and environmental components. A new literature measures specific genes and uses them as observed covariates when predicting outcomes. I will discuss these two types of work in terms of how they may inform social policy. I will argue that research on heritability is fundamentally uninformative for policy analysis, but make a cautious argument that research using genes as covariates is potentially informative.

Summary of Lecture 1

Heritability: The proportion of observed differences in a trait among individuals from a certain population that is due to genetic differences among these individuals

Answering the main questions:

Why should economists be excited about genetics?
- The three laws of behavioral genetics also holds for traits in which economists are interested

How to estimate heritability using twins?
- Exploit zygosity differences with ACE models

Is the estimation of heritability policy-relevant?
- Not directly (Goldberger), but indirectly by pointing to the presence of inequality of opportunity (Taubman)
- Longitudinal approach needed (with focus on c^2 before and after an environmental change)
- Genes! (Manski)

What comes next?

- Tutorial 1: Practice with the classical twin study in *R*
  - Go through the step-by-step tutorial (Canvas → Pages → Tutorial 1) yourself / together
  - Make a start with the Individual assignment (See end of Tutorial 1.pdf and Individual assignment.pdf on Canvas)
- Make teams for Group assignment and register them on Canvas (see Group assignment.pdf on Canvas)
  - The first presentation will be Lecture 3 (two weeks from now, *register quickly if want to be first*)
- Lecture 2
  - What are the promises of genoeconomics?
  - The basics of genetics: What is a “SNP”?
  - How can we use Mendel’s laws to understand natural selection on socio-economic traits?

Economics & Genetics Lecture 1

DR. NIELS RIETVELD, NRIETVELD@ESE.EUR.NL
OFFICE MANDEVILLE T18-29