The Changing Genetic Architecture of Health and Development within U.S. families

Dalton Conley
Princeton University & NBER
Outline of Talk

• Background motivation
• Power problems in measuring SNP effects
• Measuring $h^2$: Proxy phenotype method
• Changing genetic penetrance by birth cohort
• Phenotypic & Polygenic Similarity in U.S. Spouses
• Shifting predictors of fertility over time
• Toward GxE across and within families
• Implications
WHAT MONEY CAN'T BUY

FAMILY INCOME AND CHILDREN'S LIFE CHANCES

(1997)
The Bell Curve (1994)

- Used NLSY ’79
- Measured cognitive ability
  - Endogenous to environment
- High and rising returns to IQ
  - 1957 – 64 birth cohort
- The Bell Curve Revisited:
  - birth cohorts 1920-1960
  - genetic data
Problem of N

Duncan and Keller (2011, Am J Psychiatry); courtesy of D. Cesarini
Power Problems

• Typical sample sizes in the range $N = 100$ to $3,000$.
• Studies only well-powered under the assumption that effect sizes have fairly large $R^2$, e.g., at least 2.7% for $N = 345$.
• But average G effects are much smaller:
  – Smoking and CHRNA3: $R^2 \approx 0.5\%$.
  – BMI and FTO: $R^2 \approx 0.3\%$.
  – Largest identified SNP for height: $R^2 \approx 0.4\%$.
  – Largest identified SNP for educational attainment: $R^2 \approx 0.03\%$.
  – Largest identified SNP for subjective well-being: $R^2 \approx 0.01\%$. 
The distribution of effect sizes ($R^2$) for the 74 hits reported by Okbay et al. (2016)

$R^2 \sim .02$, sample size for 80% power: 39,240
“G”: Polygenic score (PGS) from a genome-wide association study (GWAS)

- GWAS (large meta-analysis of genetic data) identified 74 single nucleotide polymorphisms (SNPs) associated with years of schooling completed (Okbay et al., 2016)
- From GWAS, we create a PGS → single measure of genetic propensity
  - A weighted average across the number of SNPs ($n$) of the number of reference alleles ($0, 1$, or $2$) at that SNP multiplied by the score for that SNP ($\beta$):

\[
P GS_i = \sum_{j=1}^{n} (\beta_j x_{ij})
\]

<table>
<thead>
<tr>
<th>SNP</th>
<th>Reference (scored) allele</th>
<th>SNP Beta Weight (excluding HRS sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11584700</td>
<td>A</td>
<td>-0.082</td>
</tr>
<tr>
<td>rs4851266</td>
<td>T</td>
<td>0.064</td>
</tr>
<tr>
<td>rs3227</td>
<td>C</td>
<td>0.053</td>
</tr>
<tr>
<td>rs1264026</td>
<td>A</td>
<td>0.0479</td>
</tr>
</tbody>
</table>

1,411,964 SNPs were used to construct the education PGS $R^2 = 5.5\%$ (in HRS)
## SNP $h^2$ and predictive power of PGS for various phenotypes

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Paper</th>
<th>GWAS N</th>
<th>$h^2_{SNP}$</th>
<th>$R^2_{PGS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>Wood et al. (2014)</td>
<td>253,288</td>
<td>50%</td>
<td>17%</td>
</tr>
<tr>
<td>BMI</td>
<td>Locke et al. (2015)</td>
<td>339,224</td>
<td>21%</td>
<td>7%</td>
</tr>
<tr>
<td>EA</td>
<td>Rietveld et al. (2013)</td>
<td>101,069</td>
<td>22%</td>
<td>2%</td>
</tr>
<tr>
<td>EA</td>
<td>Okbay et al. (accepted)</td>
<td>394,769</td>
<td>22%</td>
<td>7%</td>
</tr>
<tr>
<td>SWB</td>
<td>Okbay et al. (forth.)</td>
<td>298,420</td>
<td>9%*</td>
<td>1%</td>
</tr>
<tr>
<td>Age at first birth</td>
<td>Barban et al. (in prep)</td>
<td>238,064</td>
<td>15%</td>
<td>1%</td>
</tr>
</tbody>
</table>

* From Rietveld et al. (2013, *PNAS*).

Table courtesy of Daniel Benjamin
Ever Larger Sample Sizes

- Education
- BMI
- Height

Out-of-Sample Predictive Power (%)

- Largest GWAS
- ~1M Discovery N
- ~2M
Dataset combines socio-demographic and genotype data from the HRS

- **Health and Retirement Study (HRS):** Biannual, multidisciplinary, nationally representative longitudinal study of individuals over the age of 50 and their spouses (1992-2016)
- In 2009 the HRS genotyped 12,507 respondents who provided DNA samples (saliva) in 2006 and 2008
  - Create survival weights to account for mortality attrition
- Link Genotype Data V.1 (2006 and 2008 samples) to 1) HRS public use data and 2) restricted date of birth files
- Restrict to non-Hispanic whites for reasons of genetic linkage structure
What is the PGS measuring?

- Captures a wide swath of genetic influences that contribute to educational attainment
- GWAS findings implicate SNPs in regions that regulate brain-specific gene expression in neural tissue as well as biological pathways that regulate forebrain development, abnormal cerebral cortex morphology, sprouting of dendrites and dendritic spine organization, neuronal signaling, and synaptic plasticity throughout the lifespan
- Pathways are myriad and ultimately not completely knowable and probably include highly contextual socially mediated mechanisms
- Summative measure of randomly-assigned genetic endowment at birth

✔ Improves upon prior estimates of endowment like IQ because it is not endogenous
How do we know it is exogenous?

PGS is normally distributed
Predictive Power of SNPs

The relationship between standardized coefficients from the first-stage regression of years of schooling on the education-associated SNPs in the Education Sample (x axis) and standardized coefficients from the second-stage regression of cognitive performance.

Cornelius A. Rietveld et al. PNAS 2014;111:13790-13794
Coefficients from regression of standardized cognitive phenotype (total word recall or total mental status) on standardized polygenic score within age category controlling for sex and clustering SEs by individual (details in SI Appendix, section 14).

Cornelius A. Rietveld et al. PNAS 2014;111:13790-13794
Proposition 1:
Social stratification is increasingly giving way to genotypic stratification

High cognitive ability as of the 1990s means, more than ever before, that the chances of success in life are good and getting better all the time, and these are *decreasingly* affected by the social environment, which by extension indicates that they must be *increasingly* affected by genes. (Pp. 109-110)
Genetics of smoking behavior

Proposition 2:
Spouses are sorting genetically / genotypically to a greater extent in recent cohorts

Add to this phenomenon known as assortative mating. Likes attract when it comes to marriage, and intelligence is one of the most important of those likes. When this propensity to mate by IQ is combined with increasingly efficient educational and occupational stratification, assortative mating by IQ has more powerful effects on the next generation than it had on the previous one. This process too seems to be getting stronger, part of the brew creating an American class system. (Pp. 91-92)
How Genetically Similar Are Spouses?

Changing Phenotypic Spousal Assortment
Changing Genotypic Spousal Assortment
Proposition 3:
Reproductive success is increasingly stratified by genotype

The professional consensus is that the United States has experienced dysgenic pressures throughout either most of the century (the optimists) or all of the century (the pessimists). Women of all races and ethnic groups follow this pattern in similar fashion. There is some evidence that blacks and Latinos are experiencing even more severe dysgenic pressures than whites, which could lead to further divergence between whites and other groups in future generations. (Pp. 341)
Changing Phenotype-NEB Relationships
Changing Genotype-NEB Relationships
Sensitivity Checks

• Mortality bias
  – Include higher order marriages
  – Compared quantiles of those who have only one surviving spouse genotyped
  – Weights adjusting for mortality before 2006 genotyping and bt 2006 - 2012
  – Rerun on Framingham Heart Study G2, median age 38

• Changing genetic architecture
  – Compared discovery samples with HRS sample.
    • All younger

• Power to detect marginal GxE effect
  • Education: 48%; Depression: 57%; Height 65%; BMI ~99.9%
  • Higher for SIMEX corrected
Sensitivity Checks

- Changing genome-wide patterns
  - GREML on younger/older split for fertility and assortative mating
    - NEB .20 v. .19 (sorry, R.A. Fisher)
    - GAM: .050 v. .045
  - Beale urban-rural split does not explain
  - PCs do not show changes in assortative mating patterns
  - PCs recalculated based on younger/older sample
  - PGS analysis done residualized on 10 PCs & 0 PCs

- Rotate demeaned data 45° to account for symmetrical heteroscedasticity
  - Regress within-family variance on between family mean PGS dispersion
  - Spouses run separately: by sex and randomly ordered
  - Reverse X and Y

- Measurement error
  - Simulation and Extrapolation Method (SIMEX) analysis
G x E as holy grail of sociogenetics
Exogenous “E”: The Vietnam Lotteries

Wave 1: December 1969
• 1944-1950 birth cohorts
• RSN’s<=195 called

Wave 2: July 1970
• 1951 birth cohort
• RSN’s<=125 called

Wave 3: August 1971
• 1952 birth cohort
• RSN’s<=95 called

Wave 4: February 1972
• 1953 birth cohort
• Never called to service

Difference in Predicted Years of Education
Veterans versus Non-Veterans
Social Genetic Effects within the Family

Niche formation, SUTVA or both?
Is the education PGS stronger within than across families?

**Effect of Genetic Score Between and Within Families**

- Within Families
- Between Families
Logic of Inquiry

• Four phenotypes / PGSs
  – Height – least environmentally biased / no niche formation
  – BMI – small environmental bias / slight niche formation
  – Education – strong environmental effects / slight niche formation
  – PPVT – strong environmental effects / moderate niche formation
  – Depression – strong environmental effects / major niche formation

• National Longitudinal Study of Adolescent to Adult Health
  – ~ 900 Sibling pairs including DZ twins: Genome-wide data
Sibling Genotype as Environment: MTFS

Effect of Standardized Education PGS of Self and Sibling on Ego's Years of Schooling

N = 397

No corresponding effect for height or BMI
Extending GxE studies

- Most GxE studies use measures of environment that are non-randomly distributed (e.g., parental divorce)
  - Potential for gene-environment correlation (rGE)
  - Does G affect response to E or affect risk of exposure
  - Possibility of epistatic interactions
  - Better to deploy known-to-be exogenous forms of environmental variation (E):
    - Birth weight differences within twin pairs
    - Policy changes: Compulsory Schooling, Draft Lottery, GI Bill, etc.
    - Societal shocks: 9/11, Great Recession, earthquakes, etc.
    - Others: Peer composition in classrooms/dorms, plant closure, etc.
  - Use polygenic scores with PCs, sib comparisons or control for parental genotype (TDT) to ensure G is not proxying E
1. As behavioral, lifestyle diseases wax in importance to population health, it is important to estimate genetic and environmental effects using representative samples.

2. More efficient and more consistent estimation of environmental variables through use of PGSs:
   - Control variables; or
   - Mendellian randomization (via Egger regression and placebo test).

3. Heritability tells us about dynastic policy effects.

4. Can reveal heterogeneous treatment effects:
   - PGS v. vPGS; can inform policy-treatment design.

5. Consumer genomics: 23andme + okCupid + IVF.
Thanks!

• Collaborators: Ben Domingue, Chris Dawes, Kathie Harris, Jason Boardman, Jason Fletcher, Dan Belsky, Thomas Laidley, Lauren Schmitz, and the SSGAC
• Datasets: HRS; FHS; MTFS; Add Health
• Funders...

John Simon Guggenheim Memorial Foundation Fellowship to Assist Research and Artistic Creation
Robert Wood Johnson Foundation
National Science Foundation
2017: Gattaca rising?
\[ R^2 \approx h_{SNPs}^2 \times (1 + \frac{M}{Nh_{SNPs}^2})^{-1} \]
Table of Consanguinity
Showing degrees of relationship
Data Structure: FHS

1948: Generation One: 5,209 men and women between the ages of 30 and 62

1971: Generation Two: 5,124 of the original participants' adult children and their spouses

G2 specimen collection: 1980s – 1990s (ages 5 – 70 in 1971)

2002: Generation Three + Two: 1573

3,514 children + 1,576 spouses
76 % Genotyped. 741 pairs with complete data

Median age: 38
Sensitivity Check: FHS G2 (Median Age 38)

<table>
<thead>
<tr>
<th></th>
<th>Phenotype b/se</th>
<th>Genotype b/se</th>
</tr>
</thead>
<tbody>
<tr>
<td>fath_education</td>
<td>0.297*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>-0.000**</td>
</tr>
<tr>
<td>age</td>
<td>-0.064</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>fath_educ_age</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>fath_score</td>
<td>0.481*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td></td>
</tr>
<tr>
<td>fath_score_age</td>
<td>-0.007</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>constant</td>
<td>10.765***</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(1.80)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>R-sqr</td>
<td>0.299</td>
<td>0.063</td>
</tr>
<tr>
<td>dfres</td>
<td>737</td>
<td>737</td>
</tr>
<tr>
<td>BIC</td>
<td>2987.7</td>
<td>-15439.5</td>
</tr>
</tbody>
</table>

* p<0.05, ** p<0.01, *** p<0.001
PGS – Pheno Controlling for 10 PCs
A Q–Q plot for a regression of cognitive performance on the education-associated SNPs (black circles) with a 95% confidence interval around the null hypothesis (gray shaded region) and a Q–Q plot for a regression of cognitive performance on the theory-based...
“E”: Vietnam-era draft lotteries

Random nature of the draft lotteries has been used as a quasi-natural experiment to identify the impact of Vietnam-era service. (e.g. Hearst, Newman, & Hulley 1986; Angrist 1990; Dobkin & Shabani 2009; Angrist & Chen 2011; Angrist, Chen & Song 2011; Conley & Heerwig 2012)

Wave 1: December 1969
- 1944-1950 birth cohorts
- RSN’s<=195 called

Wave 2: July 1970
- 1951 birth cohort
- RSN’s<=125 called

Wave 3: August 1971
- 1952 birth cohort
- RSN’s<=95 called

Wave 4: February 1972
- 1953 birth cohort
- Never called to service


Final sample: **504 white non-Hispanic men** born between 1948-1952 from the 1992-2010 HRS waves who were genotyped in 2006 or 2008
<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Non-veteran</th>
<th>Veteran</th>
<th>Diff.</th>
<th>t stat.</th>
<th>Not Draft Eligible</th>
<th>Draft Eligible</th>
<th>Diff.</th>
<th>t stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drafted</td>
<td>0.435</td>
<td>0.366</td>
<td>0.555</td>
<td>-0.188***</td>
<td>(-4.16)</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of education PGS</td>
<td>0</td>
<td>0.114</td>
<td>-0.202</td>
<td>0.315***</td>
<td>(3.43)</td>
<td>-0.025</td>
<td>0.032</td>
<td>-0.057</td>
<td>(-0.63)</td>
</tr>
<tr>
<td>Years of education</td>
<td>14.188</td>
<td>14.478</td>
<td>13.676</td>
<td>0.802***</td>
<td>(3.86)</td>
<td>14.316</td>
<td>14.023</td>
<td>0.293</td>
<td>(1.44)</td>
</tr>
<tr>
<td>Years of college (0-5)</td>
<td>2.357</td>
<td>2.665</td>
<td>1.813</td>
<td>0.851***</td>
<td>(4.86)</td>
<td>2.446</td>
<td>2.242</td>
<td>0.204</td>
<td>(1.17)</td>
</tr>
<tr>
<td>Highest degree completed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No degree</td>
<td>0.048</td>
<td>0.056</td>
<td>0.033</td>
<td>0.023</td>
<td>(1.16)</td>
<td>0.049</td>
<td>0.046</td>
<td>0.003</td>
<td>(0.18)</td>
</tr>
<tr>
<td>High school degree</td>
<td>0.49</td>
<td>0.422</td>
<td>0.61</td>
<td>-0.188***</td>
<td>(-4.10)</td>
<td>0.467</td>
<td>0.521</td>
<td>-0.054</td>
<td>(-1.20)</td>
</tr>
<tr>
<td>Associate degree</td>
<td>0.073</td>
<td>0.047</td>
<td>0.121</td>
<td>-0.074***</td>
<td>(-3.09)</td>
<td>0.081</td>
<td>0.064</td>
<td>0.017</td>
<td>(0.71)</td>
</tr>
<tr>
<td>Bachelor's degree</td>
<td>0.236</td>
<td>0.28</td>
<td>0.159</td>
<td>0.120***</td>
<td>(3.07)</td>
<td>0.232</td>
<td>0.242</td>
<td>-0.010</td>
<td>(-0.27)</td>
</tr>
<tr>
<td>Advanced degree</td>
<td>0.153</td>
<td>0.196</td>
<td>0.077</td>
<td>0.119***</td>
<td>(3.60)</td>
<td>0.172</td>
<td>0.128</td>
<td>0.0441</td>
<td>(1.36)</td>
</tr>
<tr>
<td>Ever married</td>
<td>0.869</td>
<td>0.863</td>
<td>0.879</td>
<td>-0.016</td>
<td>(-0.50)</td>
<td>0.874</td>
<td>0.863</td>
<td>0.011</td>
<td>(0.35)</td>
</tr>
<tr>
<td>N</td>
<td>504</td>
<td>322</td>
<td>182</td>
<td></td>
<td></td>
<td>285</td>
<td>219</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Education PGS is standardized on the analytic sample. *p<0.10, **p<0.05, ***p<0.01
Use IV (2SLS) approach to estimate G x E

• Treat veteran status (VET) and its interaction with the PGS as endogenous/non-random:
\[ Y_i = \alpha_1 VET_i + \alpha_2 VET_i \times PGS_i + X_i \beta + \epsilon_i \]

• Use draft eligibility (DRAFT) as an instrument for veteran status in the first stage:
\[ VET_i = \pi_1 DRAFT_i + \pi_2 DRAFT_i \times PGS_i + X_i \pi_3 + \eta_i \]
\[ VET_i \times GRS_i = \gamma_1 DRAFT_i + \gamma_2 DRAFT_i \times PGS_i + X_i \gamma_3 + \rho_i \]
IV 2SLS estimation (cont.)

• Substitute first stage equations into the structural equation:

\[
Y_i = \alpha_1[\pi_i DRAFT_i + \pi_i DRAFT_i \times PGS_i + X_i' \pi_3 + \eta_i] + \\
\alpha_2[\gamma_1 DRAFT_i + \gamma_2 DRAFT_i \times PGS_i + X_i' \gamma_3 + \rho_i] + X_i' \beta + \epsilon_i \\
= DRAFT_i[\alpha_1 \pi_1 + \alpha_2 \gamma_1] + DRAFT_i \times PGS_i[\alpha_1 \pi_2 + \alpha_2 \gamma_2] + \\
X_i' [\beta + \alpha_1 \pi_3 + \alpha_2 \gamma_3] + [\alpha_1 \eta_i + \alpha_2 \rho_i + \epsilon_i]
\]

• Derive reduced form:

\[
Y_i = \delta_1 DRAFT_i + \delta_2 DRAFT_i \times PGS_i + X_i' \delta_3 + \zeta_i
\]
Military service reduced schooling of veterans with below average PGSs

<table>
<thead>
<tr>
<th></th>
<th>Years of education</th>
<th>Years of college</th>
<th>GED/HS degree +</th>
<th>Associate degree +</th>
<th>Bachelor's degree +</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGS = 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.055</td>
<td>-0.623</td>
<td>0.083</td>
<td>-0.108</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>(1.589)</td>
<td>(1.302)</td>
<td>(0.139)</td>
<td>(0.348)</td>
<td>(0.282)</td>
</tr>
<tr>
<td>PGS = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.736</td>
<td>-1.314</td>
<td>0.0792</td>
<td>-0.308</td>
<td>-0.125</td>
</tr>
<tr>
<td></td>
<td>(1.446)</td>
<td>(1.185)</td>
<td>(0.125)</td>
<td>(0.315)</td>
<td>(0.289)</td>
</tr>
<tr>
<td>PGS = -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-3.100**</td>
<td>-2.698**</td>
<td>0.072</td>
<td>-0.707**</td>
<td>-0.295</td>
</tr>
<tr>
<td></td>
<td>(1.464)</td>
<td>(1.200)</td>
<td>(0.122)</td>
<td>(0.315)</td>
<td>(0.373)</td>
</tr>
<tr>
<td>PGS = -2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-3.782**</td>
<td>-3.390**</td>
<td>0.069</td>
<td>-0.906***</td>
<td>-0.380</td>
</tr>
<tr>
<td></td>
<td>(1.621)</td>
<td>(1.331)</td>
<td>(0.134)</td>
<td>(0.348)</td>
<td>(0.438)</td>
</tr>
<tr>
<td>N</td>
<td>504</td>
<td>504</td>
<td>504</td>
<td>480</td>
<td>233</td>
</tr>
</tbody>
</table>
Genetics of educational attainment

• Unmeasured biological variation biases models that seek to isolate the causal effect of the social environment on educational attainment (“ability bias”)
• Past studies → used measures of IQ or scores on cognitive tests as a proxy
  – This approach has several problems:
    ✓ It doesn’t measure non-cognitive traits like personality
    ✓ Both cognitive and non-cognitive traits are affected by the social environment (e.g. IQ can be strongly affected by the uterine environment)
• Twin or sibling models can factor out genetic effects by controlling for the family environment, but this approach doesn’t give us a direct measure of genotype to deploy in observational G x E studies
Assumptions: $R^2_E = 20\%$ and $R^2_G = 0.5\%$. Source: Duncan and Keller (2011, Am J Psychiatry); courtesy of D. Cesarini.
|                                | Coef. | Std. Err. | t     | P>|t|   | [95% Conf. Interval]          |
|--------------------------------|-------|-----------|-------|-------|-----------------------------|
| ht_20_cm                       |       |           |       |       |                             |
| score_ht_std                   | 2.950194 | .7114001 | 4.15  | 0.000 | 1.551065 - 4.349323         |
| score_ht_std_B                 | -0.6300522 | .7262653 | -0.87 | 0.386 | -2.058417 - 0.7983128       |
| score_ht_std_AB                | 0.2738463 | .2350188 | 1.17  | 0.245 | -0.1883714 - 0.736064       |
| _cons                          | 171.2312 | .4927549  | 347.50| 0.000 | 170.2621 - 172.2003         |

```
.reg ht_20_cm score_ht_std score_ht_std_B
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 356</th>
</tr>
</thead>
<tbody>
<tr>
<td>F(2, 353)</td>
<td>12.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>1887.91095</td>
<td>2</td>
<td>943.955473</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>26029.4361</td>
<td>353</td>
<td>73.7377793</td>
<td>0.0676</td>
</tr>
<tr>
<td>Total</td>
<td>27917.347</td>
<td>355</td>
<td>78.6404141</td>
<td>0.0623</td>
</tr>
</tbody>
</table>

```
.reg ht_20_cm score_ht_std score_ht_std_B
```

|                                | Coef. | Std. Err. | t     | P>|t|   | [95% Conf. Interval]          |
|--------------------------------|-------|-----------|-------|-------|-----------------------------|
| ht_20_cm                       |       |           |       |       |                             |
| score_ht_std                   | 2.787876 | .6979812 | 3.99  | 0.000 | 1.415152 - 4.1606           |
| score_ht_std_B                 | -0.7681241 | .7168964 | -1.07 | 0.285 | -2.178049 - 0.6418011       |
| _cons                          | 171.451 | .4554309  | 376.46| 0.000 | 170.5553 - 172.3467         |

```
.reg ht_20_cm score_ht_std score_ht_std_B
```
. reg EDYRS score_ed2B_std score_ed2_std edscore2_AB_std

Source | SS   df    MS          Number of obs =  397
---------+----------------------------------- F(3, 393)  =  22.76
Model | 362.719185   3   120.906395  Prob > F  =  0.0000
Residual | 2087.36646  393  5.31136503  R-squared  =  0.1480
---------+----------------------------------- Adj R-squared  =  0.1415
Total | 2450.08564  396  6.18708496  Root MSE  =  2.3046

-------------------------------------------------------------
      EDYRS |  Coef.  Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------
  score_ed2B_std |   .4195989  .1425165    2.94   0.003     .1394088    .699789
  score_ed2_std |   .6321174  .1474340    4.29   0.000     .3422595    .9219754
  edscore2_AB_std |  -.1620121  .0980825    -1.65   0.099    -.3548441    .0308198
     _cons |   15.52633   .1298340   119.59  0.000    15.27107    15.78159
-------------------------------------------------------------

. reg EDYRS score_ed2B_std score_ed2_std

Source | SS   df    MS          Number of obs =  397
---------+----------------------------------- F(2, 394)  =  32.64
Model | 348.227515   2   174.113757  Prob > F  =  0.0000
Residual | 2101.85813  394  5.3346653  R-squared  =  0.1421
---------+----------------------------------- Adj R-squared  =  0.1378
Total | 2450.08564  396  6.18708496  Root MSE  =  2.3097

-------------------------------------------------------------
      EDYRS |  Coef.  Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------
  score_ed2B_std |   .400085   .1425165    2.81   0.005     .1202496    .6799205
  score_ed2_std |   .6433471  .1474340    4.36   0.000     .3531653    .9335288
     _cons |   15.52633   .1298340   119.59  0.000    15.27107    15.78159
-------------------------------------------------------------
How do we know it’s exogenous?
How do we know it’s exogenous?
Fourth “Law” as a Unifying Principle

“A typical human behavioral trait is associated with very many genetic variants, each of which accounts for a very small percentage of the behavioral variability.”


$R^2 \sim .02$, sample size for 80% power: 39,240