GENES AND SOCIOECONOMIC OUTCOMES

Steven Durlauf, University of Chicago

A Celebration of Peter Phillips’ Forty Years at Yale

October 19-20, 2018
IDENTIFICATION OF GENETIC INFLUENCES ON SOCIOECONOMIC OUTCOMES VIA KINSHIP RELATIONS

Steven Durlauf, University of Chicago

Andros Kourtellos, University of Cyprus

Chih Ming Tan, University of North Dakota
This paper explores the identification of genetic influences (genotype) on socioeconomic behavior (phenotype).

Deep controversies within social science about quantitative claims about the role of genes.

It has been extraordinary difficult to measure the role of genes vs. the environment.

We propose a constructive strategy to measure genetic influences.
• We focus on twins data.

• Twins studies are fundamental in behavioral genetics.

• The basic strategy in this literature is based on comparing correlations in outcomes of monozygotic (identical) versus dizygotic (fraternal) twins to infer the variance contribution of genes to outcomes.

• This approach has been severely criticized by Arthur Goldberger among others.
Goldberger’s critique was based on two identification assumptions:

1. uncorrelatedness of genetic and environmental factors
2. equal similarity of family environment for dizygotic twins as for monozygotic twins

Both assumptions are problematic if one thinks of parents as choosing the family environment for a child.
There are two different methodologies

- **kinship studies:**
  - genome is latent
  - exploits genetic overlaps between twins, siblings, cousins, etc.
  - second moments

- **genomic studies:**
  - genome is observed
  - exploits DNA data and in particular 10 million single nucleotide polymorphism (SNPs) variables
  - first moments
  - big data methods

- these two different strategies ought to be unified into a common framework
Introduction

We treat the genotype of an individual as latent variable in a particular economic environment.

In particular, we propose to model the joint evolution of genetic and environmental influences across generations by developing a behavioral model of socioeconomic outcomes that integrates

- the classic family investment model of intergenerational mobility due to Becker and Tomes (1979) and Loury (1981).

We demonstrate how economic theory places restrictions on kinship correlations that are sufficient to identify the role of genetic factors.

We provide a different way to ask the question about the relationship between genes environment and socio-economic outcomes.

In doing so we challenge the conventional approach by proposing a new way to measure the role of genes.

There’s an interplay between the genotype the environment and that interplay leads to a certain set of socio-economic outcomes.
• Variance decomposition in a temporal context.

• Variance contributions of innovations in genes using a SVAR type model.

• The idea is that a parental vector of capabilities is transmitted to the capabilities of child and in turn these capabilities joined with incentives and social environment determine child outcomes.
Our strategy applies to general kinship studies.

So does the Goldberger critique, as we will demonstrate.

A message of our research program is that, without a priori assumptions, identification fails. There are no clever IV strategies, etc.
In our context gene-environment correlation is derived from an economic model.

We deviate from Goldberger in the sense that economic theory will restrict and provide bounds.

We exploit the nonlinear structure of our model.
**Outline of the Talk**

1. Development of Impossibility Theorems for statistical models of genes, environment, outcomes (generalized Goldberger).

2. Description of behavioral model which produces appropriate and identified measure of role of genes.
The canonical model for nature-nurture decompositions is given by the ACE model for twins $i$ and $i'$ in family $j$

$$\omega_{i,j,t} = aA_{i,j,t} + cC_{i,j,t} + eE_{i,j,t}, \quad (1)$$

- $\omega_{i,j,t}$: outcome of offspring $i$ in generation $t$ (phenotype),
- $A_{i,j,t}$: genetic component (genotype),
- $C_{i,j,t}$: shared environment (family),
- $E_{i,j,t}$: nonshared environment (luck).
• The objective of twins studies is to identify the contributions of the three factors to overall variance, in particular the genetic coefficient, $a$, since this coefficient is the basis for measuring the roles of nature versus nurture.

• To do this, consider the following three assumptions.
Assumption A.1

The different determinants of $\omega_{i,j,t}$ are uncorrelated with one another, i.e.

$$\text{Cov}(A_{i,j,t}, C_{i,j,t}) = \text{Cov}(A_{i,j,t}, E_{i,j,t}) = \text{Cov}(C_{i,j,t}, E_{i,j,t}) = 0$$  \hspace{1cm} (2)

This renders a variance decomposition meaningful.
Assumption A.2

Covariances of components of outcomes between twins are determined by combination of assumptions and laws of genetics.

(i) Parallel to (2), interactions between covariances across twins are zero.

\[ \text{Cov}(A_{i,j,t}, C_{i',j,t}) = \text{Cov}(A_{i,j,t}, E_{i',j,t}) = \text{Cov}(C_{i,j,t}, E_{i',j,t}) = 0 \text{ if } i \neq i' \]  

(ii) Twins in same house experience are equally correlated in their exposure to environment regardless of twin type: monozygotic (m) or dizygotic (d).

\[ \text{Cov}(C_{i,j,t}, C_{i',j,t}|m) = \text{Cov}(C_{i,j,t}, C_{i',j,t}|d) = 1 \text{ if } i \neq i' \]
Assumption A.3

Genetic covariation is determined by twin type.

\[
\text{Cov}(A_{i,j,t}, A_{i',j,t}) = \begin{cases} 
1, & \text{if } i \neq i', (i, i') \text{ monozygotic} \\
0.5, & \text{if } i \neq i', (i, i') \text{ dizygotic}
\end{cases}
\]  

(5)
• This leads to a system of 3 equations with three moments (variances of outcomes for individuals and covariances by twin type) and 3 unknown parameters, so they are identified.

\[
\text{Var}(\omega_{i,j,t}|m) = \text{Var}(\omega_{i,j,t}|d) = a^2 + c^2 + e^2
\] (6)

\[
\text{Cov}(\omega_{i,j,t}, \omega_{i',j,t}|m) = a^2 + c^2
\] (7)

\[
\text{Cov}(\omega_{i,j,t}, \omega_{i',j,t}|d) = 0.5a^2 + c^2
\] (8)

• These identify the variance components.
STANDARD ACE

- The genetic component:

\[ a^2 = 2 \left( \text{Cov}(\omega_{i,j,t}, \omega_{i',j,t} | m) - \text{Cov}(\omega_{i,j,t}, \omega_{i',j,t} | d) \right) \]  (9)

- The shared environment:

\[ c^2 = 2\text{Cov}(\omega_{i,j,t}, \omega_{i',j,t} | d) - \text{Cov}(\omega_{i,j,t}, \omega_{i',j,t} | m) \]  (10)

- The nonshared environment:

\[ e^2 = \text{Var}(\omega_{i,j,t} | m) - \text{Cov}(\omega_{i,j,t}, \omega_{i',j,t} | m) \]  (11)
In a long series of papers, Arthur Goldberger systematically critiqued twins studies.

His work, in addition to careful deconstruction of specific studies, made a general methodological criticism in that he challenged

1. two uncorrelatedness of genes and environment.

\[ \text{Cov}(A_{i,j,t}, C_{i,j,t}) \neq 0; \text{Cov}(A_{i,j,t}, C_{i,j,t}|m) \neq \text{Cov}(A_{i,j,t}, C_{i,j,t}|d) \]

2. and the equal environment assumption for monozygotic and dizygotic twins

\[ \text{Cov}(C_{i,j,t}, C_{i',j,t}|m) \neq \text{Cov}(C_{i,j,t}, C_{i',j,t}|d) \]
This leads to a system of 4 equations and 8 unknowns, so identification fails:

\[ \text{Var}(\omega_{i,j,t} | m) = a^2 + c^2 + e^2 + 2ac\text{Cov}(A_{i,j,t}, C_{i,j,t} | m) \] (12)

\[ \text{Var}(\omega_{i,j,t} | d) = a^2 + c^2 + e^2 + 2ac\text{Cov}(A_{i,j,t}, C_{i,j,t} | d) \] (13)

\[ \text{Cov}(\omega_{i,j,t}, \omega_{i',j,t} | m) = a^2 + c^2 \text{Cov}(C_{i,j,t}, C_{i',j,t} | m) + 2ac\text{Cov}(A_{i,j,t}, C_{i',j,t} | m) \] (14)

\[ \text{Cov}(\omega_{i,j,t}, \omega_{i',j,t} | d) = 0.5a^2 + c^2 \text{Cov}(C_{i,j,t}, C_{i',j,t} | d) + 2ac\text{Cov}(A_{i,j,t}, C_{i',j,t} | d) \] (15)
GOLDBERGER CRITIQUE

- Note that, in isolation, gene environment interaction leads to partial identification.
- Suppose that one allows for gene environment correlation
  \[
  \text{Cov}(A_{i,j,t}, C_{i,j,t} | m) = \text{Cov}(A_{i,j,t}, C_{i,j,t} | m) = r
  \]
  but preserves the identical environment for twins assumption. i.e.,
  \[
  \text{Cov}(C_{i,j,t}, C'_{i',j,t} | m) = \text{Cov}(C_{i,j,t}, C'_{i',j,t} | d) = 1
  \]
- In this case, the system of second moments is
Goldberger Critique

\[
\text{Var}(\omega_{i,j,t}|m) = a^2 + c^2 + e^2 + 2acr
\]  \hspace{1cm} (16)

\[
\text{Cov}(\omega_{i,j,t}, \omega_{i',j,t}|m) = a^2 + c^2 + 2acr
\]  \hspace{1cm} (17)

\[
\text{Cov}(\omega_{i,j,t}, \omega_{i',j,t}|d) = 0.5a^2 + c^2 + 2acr
\]  \hspace{1cm} (18)

- While there are 3 equations in 4 unknowns \((a^2, c^2, e^2, r)\).
- The nonlinearity in the system means that \(a = 0\) is testable, since if it holds,

\[
\text{Cov}(\omega_{i,j,t}, \omega_{i',j,t}|m) = \text{Cov}(\omega_{i,j,t}, \omega_{i',j,t}|d)
\]
Generalized Kinship Models

- Within behavioral genetics, twins models have been extended. One popular variant is children of twins models.

- There is an impossibility theorem for identification across kinship groups that can be based on Goldberger’s approach to twins.

- Suppose there are K distinct pairs. In the case of twins, there are two: monozygotic and dizygotic. In the case of children of twins, there are two types of cousins, children of monozygotic and children of dizygotic.
**Generalized Kinship Models**

\[ \Omega_{k1} = E \left[ (A_{i,j,t}, C_{i,j,t}, E_{i,j,t})(A_{i,j,t}, C_{i,j,t}, E_{i,j,t})' | k_1 \right] \]  \hspace{1cm} (19)

\[ \Omega_{k2} = E \left[ (A_{i,j,t}, C_{i,j,t}, E_{i,j,t})(A'_{i',j,t}, C'_{i',j,t}, E'_{i',j,t})' | k_2 \right] \]  \hspace{1cm} (20)

\[ \text{Var}(\omega_{i,j,t}| k_1) = (a, c, e)\Omega_{k1}(a, c, e)' \]  \hspace{1cm} (21)

\[ \text{Cov}(\omega_{i,j,t}, \omega_{i',j,t}| k_2) = (a, c, e)\Omega_{k2}(a, c, e)' \]  \hspace{1cm} (22)

- This means that every kinship pair type that is added to a set of initial kinship variances and covariances introduces 12 unknown elements of \( \Omega_{k1} \) and \( \Omega_{k2} \).
Generalized Kinship Models

Impossibility Theorem #1

Coefficients \((a, c, e)\) are not identified for any collection of kinship types without restrictions on \(\Omega_{k1}\) and \(\Omega_{k2}\)
GENERALIZED KINSHIP MODELS

(Relatively) Credible Restrictions

Assumption B.1:

Restriction on nonshared environment, which is treated by many studies as luck:

\[ \text{Cov}(A_{i,j,t}, E_{i,j,t}|k) = \text{Cov}(C_{i,j,t}, E_{i,j,t}|k) = 0 \]

\[ \text{Cov}(A_{i,j,t}, E'_{i',j,t}|k) = \text{Cov}(C_{i,j,t}, E'_{i',j,t}|k) = 0 \]

and

\[ \text{Var}(E_{i,j,t}|k) \text{ independent of } k \]

• This means nonshared environment is uncorrelated with other factors and its variance is the same for all population types.

• Not clear that this is sensible in family models.
Assumption B.2:

Restriction on distribution of genes across kinship types

\[ \text{Var}(A_{i,j,t} | k) \text{ constant across } k \]

- This rules out differences in mating patterns.
**Assumption B.3:**

Restriction on distribution of environment across kinship types

\[ \text{Var}(C_{i,j,t}|k) \text{ constant across } k \]

It is not clear if this assumption can hold if shared environment is endogenous.
Even with these assumptions, $\text{Cov}(A_{i,j,t}, C_{i,j,t}|k)$ and $\text{Cov}(A_{i,j,t}, C_{i',j,t}|k)$ appear in each variance covariance relationship for kinship type $k$.

As a result, increases in observable second moments by addition of new types of kinship types, cannot decrease the number of unknown parameters and moments.
Generalized Kinship Models

Impossibility Theorem 2:

Conditional on Assumptions B.1, B.2, B.3, no set of kinship relationships can identify \((a, c, e)\).

- This is the generalized Goldberger critique.
A second approach to measuring genetic influences is based on intergenerational dynamics.

Note that this analysis is distinct from the intragenerational analysis based on the genetic simplex model that was proposed to investigate the covariance structure of longitudinal phenotypic measures in twin studies.
• The evolution of the components of kinship models can be written as

\[
\begin{bmatrix}
A_{i,j,t} \\
C_{i,j,t} \\
E_{i,j,t}
\end{bmatrix}
\begin{bmatrix}
X_{i,j,t}
\end{bmatrix}
= \begin{bmatrix}
\phi_{AA} & \phi_{AC} & \phi_{AE} \\
\phi_{CA} & \phi_{CC} & \phi_{CE} \\
\phi_{EA} & \phi_{EC} & \phi_{EE}
\end{bmatrix}
\begin{bmatrix}
A_{i,j,t-1} \\
C_{i,j,t-1} \\
E_{i,j,t-1}
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{A,i,j,t} \\
\varepsilon_{C,i,j,t} \\
\varepsilon_{E,i,j,t}
\end{bmatrix}
\begin{bmatrix}
X_{i,j,t-1}
\end{bmatrix}
\] (23)
We assume a common factor representation for the errors using two types of shocks: common environment shocks and twin-specific idiosyncratic structural shocks.

\[
\begin{bmatrix}
\varepsilon_{A,j,t} \\
\varepsilon_{C,j,t} \\
\varepsilon_{E,j,t}
\end{bmatrix}
= \begin{bmatrix}
\lambda_{AA} & \lambda_{AC} & \lambda_{AE} \\
\lambda_{CA} & \lambda_{CC} & \lambda_{CE} \\
\lambda_{EA} & \lambda_{EC} & \lambda_{EE}
\end{bmatrix}
\begin{bmatrix}
\bar{\varepsilon}_{A,j,t} \\
\bar{\varepsilon}_{C,j,t} \\
\bar{\varepsilon}_{E,j,t}
\end{bmatrix}
+ \begin{bmatrix}
\psi_{AA} & \psi_{AC} & \psi_{AE} \\
\psi_{CA} & \psi_{CC} & \psi_{CE} \\
\psi_{EA} & \psi_{EC} & \psi_{EE}
\end{bmatrix}
\begin{bmatrix}
\bar{\tilde{\varepsilon}}_{A,i,j,t} \\
\bar{\tilde{\varepsilon}}_{C,i,j,t} \\
\bar{\tilde{\varepsilon}}_{E,i,j,t}
\end{bmatrix}
\]

These two types of shocks are assumed to be mutually orthogonal and innovations for all \(i, j,\) and \(t.\)

\(\lambda\)'s and \(\psi\)'s place restrictions on the contemporaneous relationship of the shocks.
This approach allows for a different conceptualization of the role of genes.

First, thought experiments do not involve nature versus nurture. Rather, one can think of variance decompositions that lead to the evaluation of the contribution of $Var(\varepsilon_{A,i,j,t})$ to $Var(\omega_{i,j,t})$.

Second, moments relating observables of parents and offspring are relevant for identification questions.
Dynamic Models

- The evolution of the outcome of interest is defined by the AR (1)-ACE

\[
\omega_{i,j,t} = \left[ a\phi_{AA} + c\phi_{AC} \quad a\phi_{CA} + c\phi_{CC} \quad a\phi_{AE} + c\phi_{CE} \right] \begin{bmatrix}
A_{i,j,t} \\
C_{i,j,t} \\
E_{i,j,t}
\end{bmatrix}
+ \begin{bmatrix}
A \\
C \\
E
\end{bmatrix} \begin{bmatrix}
\varepsilon_{A,i,j,t} \\
\varepsilon_{C,i,j,t} \\
\varepsilon_{E,i,j,t}
\end{bmatrix}
\]

(24)

- Large number of parameters implied by kinship relations.
The conventional AR(1) IGE regression is

$$\omega_{i,j,t} = \beta \omega_{i,j,t-1} + \psi_{i,j,t}$$  \hspace{1cm} (25)

The IGE parameter $\beta$ (identified of course since (25) is a projection) contains information on role of genes.
The AR(1) ACE and AR(1) IGE model are not consistent with one another outside of nongeneric cases.

\[
\begin{align*}
\omega_{i,j,t} = (a & \ c & e) \begin{bmatrix}
A_{i,j,t} \\
C_{i,j,t} \\
E_{i,j,t}
\end{bmatrix} \\
= \begin{bmatrix}
\phi_{AA} & \phi_{AC} & \phi_{AE} \\
\phi_{CA} & \phi_{CC} & \phi_{CE} \\
0 & 0 & 0
\end{bmatrix} \\
\begin{bmatrix}
A_{i,j,t-1} \\
C_{i,j,t-1} \\
E_{i,j,t-1}
\end{bmatrix} + \begin{bmatrix}
a & c & e
\end{bmatrix} \begin{bmatrix}
\varepsilon_{A,i,j,t} \\
\varepsilon_{C,i,j,t} \\
\varepsilon_{E,i,j,t}
\end{bmatrix}
\end{align*}
\]

\[
= \begin{bmatrix} a\phi_{AA} + c\phi_{AC} & a\phi_{CA} + c\phi_{CC} & a\phi_{AE} + c\phi_{CE} \end{bmatrix} \begin{bmatrix} A_{i,j,t-1} \\
C_{i,j,t-1} \\
E_{i,j,t-1}
\end{bmatrix} + \begin{bmatrix} a & c & e \end{bmatrix} \begin{bmatrix} \varepsilon_{A,i,j,t} \\
\varepsilon_{C,i,j,t} \\
\varepsilon_{E,i,j,t}
\end{bmatrix}
\]

(26)
Dynamic Models

• Which makes clear that an AR(1) formulation for IGE implies that

\[
\begin{bmatrix}
    a\phi_{AA} + c\phi_{AC} & a\phi_{CA} + c\phi_{CC} & a\phi_{AE} + c\phi_{CE}
\end{bmatrix}
\begin{bmatrix}
    A_{i,j,t-1} \\
    C_{i,j,t-1} \\
    E_{i,j,t-1}
\end{bmatrix} = 
\begin{bmatrix}
    A_{i,j,t-1} \\
    C_{i,j,t-1} \\
    E_{i,j,t-1}
\end{bmatrix} = 
\beta \omega_{i,j,t-1}
\]

• This last formulation indicates that equivalence is non-generic. Intuition is simple: aggregation does not preserve AR(1) property of components.
This is obvious for the special case in which all nondiagonal elements of AR(1) coefficient matrix are zero. For this case, consistency requires

\[ \phi_{AA} a = \beta a \]
\[ \phi_{CC} c = \beta c \]

which implies that \( \phi_{AA} = \phi_{CC} \) if \( a, c \neq 0 \).

This is really just a restatement of the fact that a sum of AR(1) processes is not AR(1).

This point is discussed by Becker and Tomes (1986) and Solon (1993).
A DIFFERENT APPROACH TO MEASURING GENETIC INFLUENCES

• The intergenerational structure produces a form of a vector autoregression.

• Variance decompositions, of course, standard tool in that methodology.

• For genes and kinship, the approach is applicable and suggests a different way to think about the roles of genes, one that eliminates any genes/environment orthogonality.
A Different Approach to Measuring Genetic Influences

Impossibility Theorem 3 (trivial)

Autocovariance function of family incomes does not identify genetic contributions, even if variance/covariance structure of shocks is given.
A Different Approach to Measuring Genetic Influences

Impossibility Theorem 4 (slightly less trivial)

Autocovariance function of family incomes with twins does not identify genetic contributions, even if variance covariance structure of shocks is given.
• A two-period overlapping generations model.

• There are $N$ dynasties indexed by $j$ in this economy.

• Each individual born at time $t$ lives for two periods; i.e., childhood ($c$) and adulthood ($a$).

• Individuals consume and receive parental investments during childhood, they find employment and produce exactly two children during adulthood. Specifically, each adult cohort of dynasty $j$ produces a pair of twins, indexed by $i = 1, 2$. 
• In our context gene-environment correlation is derived from an economic model.

• We deviate from Goldberger in the sense that economic theory will restrict and provide bounds.

• We exploit the nonlinear structure of our model.
• Variance decomposition in a temporal context.

• Variance contributions of innovations in genes using a SVAR type model.

• The idea is that a parental vector of capabilities is transmitted to the capabilities of child and in turn these capabilities joined with incentives and social environment determine child outcomes.
**Behavioral Model**

- Implied Markov process and a vector of unobserved heterogeneity:
  - shocks due to labor market luck (unshared environment)
  - shocks due to unobserved parental quality (shared environment)
  - shocks due to genetic heterogeneity

- The offspring income is a weighted sum of the shocks across periods.
• What will happen to the variance in income of the offspring if we switch off the shocks due to unobserved parental quality?

• What will happen to the variance in income of the offspring if we switch off the variation in genes?

• What is the role of the accumulation of each shock to the overall variance?
• Genes vs. Environment → Variance Contributions of Innovations in Genes vs. Environment.

• Statistical Assumptions → Behavioral Assumptions

• Atheoretical Approach → Structural Approach
GENETIC MECHANISM AND SKILLS DEVELOPMENT

- Each individual is characterized by a latent index of capabilities.

- Suppose we observe proxies of this index at two distinct points of time:
  - during the childhood period; i.e., at birth (initial endowment), $\theta_{i,j,c,0,t}$,
  - during early life of the offspring, $\theta_{i,j,c,t}$.

- $\theta_{i,j,c,0,t}$ and $\theta_{i,j,c,t}$ are latent but there exist many measurements on them that can be used to provide estimates of these phenotypic indices.
Genetic mechanism and skills development

Initial endowment:

\[ \log(\theta_{i,j,c,0,t}) = k_g g_{i,j,t} + \epsilon_{i,j,c,0,t}^{\theta_0}, \]  

(28)

- We assume that conditional on the offspring’s genotype, the initial phenotype (at birth), \( \theta_{i,j,c,0,t} \) does not depend on the genotype of parents or siblings or any other shared or unshared environmental factors.
- For now assume that \( k_g = 1 \)
GENETIC MECHANISM AND SKILLS DEVELOPMENT

The genotype equation:

\[ g_{i,j,t} = 0.5 g_{i,j,t-1} + v_{i,j,t}^g, \]  
\[ v_{i,j,t}^g = \sqrt{\rho_{A,i}^k \eta_{j,t}^g + \varepsilon_{i,j,t}^g}, \]  

- The coefficient 0.5 can be justified in the context of the classical biometric model with two alleles assuming random mating, which is free from mutation, migration, and natural selection, so that the Hardy-Weinberg equilibrium holds. \( g \) is the number of copies of a particular allele in the genotype.
- \( \rho_{A,k}, k = m, d \) is the correlation coefficient between the additive genetic factors of twins \( i' \) and \( i \) of family \( j \).
- The genetic shock \( v_{i,j,t}^g \) is correlated between the twins but uncorrelated across families \( j \) and generations: \( \rho_{A,m} = 1 \) and \( \rho_{A,d} = 0.5 \).
GENETIC MECHANISM AND SKILLS DEVELOPMENT

Phenotype at early life:

\[
\log(\theta_{i,j,c,t}) = \pi_l \log(I_{i,j,c,t}) + \pi_\theta \log(\theta_{i,j,c,0,t}) + \zeta_\theta \log(\theta_{i,j,c,0,t}) \log(I_{i,j,c,t}) \\
+ \pi_Y \log(Y_{i,j,a,t}) + \eta_{j,c,t} + \varepsilon_{i,j,c,t}
\]  

(31)

- \(I_{i,j,c,t}\) denotes early parental investments,
- \(Y_{i,j,a,t}\) denotes parental income (common to the siblings).
- The shocks \(\eta_{j,c,t}\) and \(\varepsilon_{i,j,c,t}\), respectively, denote the common family shock and idiosyncratic shock realized during early life.
Human capital process:

\[
H_{i,j,c,t} = \rho_{\theta} \log(\theta_{i,j,c,t}) + \rho_{\gamma} \log(Y_{i,j,a,t}) + \eta^H_{j,c,t} + \varepsilon^H_{i,j,c,t}
\]  

(32)

- Human capital \((H_{i,j,c,t})\) accumulation occurs at the childhood \((c)\) stage after abilities (phenotype) are realized.

- The shocks \(\eta^H_{j,c,t}\) and \(\varepsilon^H_{i,j,c,t}\) denote, respectively, an environment shock (e.g., because of family environment) that is common to both siblings and an idiosyncratic shock.
Income process

\[
\log(Y_{i,j,a,t+1}) = a_Y + wH_{i,j,c,t} + \varepsilon^Y_{i,j,a,t+1}
\]  

(33)

- \( w \) is the return to education

- the shock \( \varepsilon^Y_{i,j,a,t} \) is “labor market luck”.
Combining (32) and (33) we get

\[ \log(Y_{i,j,a,t+1}) = a_Y + wp_\theta \log(\theta_{i,j,c,t}) + wp_Y \log(Y_{i,j,a,t}) + w\eta_{j,c,t}^H + w\epsilon_{i,j,c,t}^H + \epsilon_{i,j,a,t+1}^Y \]

(34)
Further, combining (34) and (31) we get

$$
\log(Y_{i,a,t+1}) = a(\theta_{i,j,c,0,t}) + b(\theta_{i,j,c,0,t}) \log(l_{i,t})
+ w(\rho Y + \rho_o \pi Y) \log(Y_{i,j,a,t}) + v_{i,j,t+1}
$$

(35)

where

$$
a(\theta_{i,j,c,0,t}) = a_Y + w\rho_o \pi_o \log(\theta_{i,j,c,0,t}),
$$

(36)

$$
b(\theta_{i,j,c,0,t}) = w\rho_o \pi_l + w\rho_o \zeta_l \log(\theta_{i,j,c,0,t})
$$

(37)

$$
v_{i,j,t+1} = [w\rho_o \eta_{i,j,c,t} + w\eta_{i,j,c,t}] + [w\xi_{i,j,c,t} + w\rho_o \xi_{i,j,c,t} + \xi_{i,j,a,t+1}]
$$

(38)
Parent’s Decision Problem

Each parent (i.e., Adult) $i$ in dynasty $j$ born at time $t-1$ maximizes the following Cobb-Douglas utility function:

$$\max_{i,j,a,t, l_i,j,c,t, l_i',j,c,t} E \left[ (1 - \gamma_y) \log(C_{i,j,a,t}) + \gamma_y \left( (\log(Y_{i,j,a,t+1}) + (\log(Y_{i',j,a,t+1}))^{\sigma} \right)^{1/\sigma} \right]$$

subject to the budget constraint

$$Y_{i,j,a,t} = C_{i,j,a,t} + l_{i,j,c,t} + l_{i',j,c,t}$$

We assume that $\sigma = 1$ so that child utilities are perfect substitutes.
As pointed out by Heckman and Mosso (2014), parents with multiple children (twins in this case) will choose to reinforce or compensate for the initial disadvantages of their weaker child depending on the curvature of the human capital production function.

\[
\frac{l_{i,j,c,t}^*}{l_{i',j,c,t}^*} = \frac{b(\theta_i,j,c,0,t)}{b(\theta_{i'},j,c,0,t)}, \quad (41)
\]

which implies that

\[
l_{i,j,c,t}^* = \frac{\gamma_y b(\theta_i,j,c,0,t)}{1 - \gamma_y (1 - b(\theta_i,j,c,0,t) - b(\theta_{i'},j,c,0,t))} \cdot Y_{i,a,t} \equiv \psi_1(\theta_i,j,c,0,t, \theta_{i'},j,c,0,t) Y_{i,a,t} \quad (42)
\]

\[
l_{i',j,c,t}^* = \frac{\gamma_y b(\theta_{i'},j,c,0,t)}{1 - \gamma_y (1 - b(\theta_i,j,c,0,t) - b(\theta_{i'},j,c,0,t))} \cdot Y_{i',a,t} \equiv \psi_2(\theta_i,j,c,0,t, \theta_{i'},j,c,0,t) Y_{i',a,t} \quad (43)
\]
EQUILIBRIUM LAW OF MOTION

\[ \log(Y_{i,a,t+1}) = \alpha^*(\theta_{i,j,c,0,t}, \theta'_{i',j,c,0,t}) + \beta^*(\theta_{i,j,c,0,t}) \log(Y_{i,j,a,t}) + v_{i,j,t+1} \]  

(44)

\[ \alpha^*(\theta_{i,j,c,0,t}, \theta'_{i',j,c,0,t}) = a_Y + w_{\rho \theta} \pi_Y \log(\theta_{i,j,c,0,t}) + (w_{\rho \theta} \pi_l + w_{\rho \theta} \zeta_{\theta} \log(\theta_{i,j,c,0,t})) \log(\psi_i(\theta_{i,j,c,0,t}, \theta'_{i',0,t})), \]  

(45)

\[ \beta^*(\theta_{i,j,c,0,t}) = w_{\rho Y} + w_{\rho \theta} \pi_Y + w_{\rho \theta} \pi_l + w_{\rho \theta} \zeta_{\theta} \log(\theta_{i,j,c,0,t}), \]  

(46)

\[ v_{i,j,t+1} = w_{\eta^H} + w_{\varepsilon^H} + \varepsilon_{i,j,a,t+1} + w_{\rho \theta} \eta_t + w_{\rho \theta} \varepsilon_t \]  

(47)

- The IGE depends on the initial phenotype and therefore depends on genes.
- Similarly, the intercept and therefore the long-run average of inter-generational incomes also depends on the initial phenotype.
Special Case: Becker-Tomes

- The model above nests the pure family investment model (Becker and Tomes (1979, 1986), Solon (1992)) when we switch off the interaction between parental investments and the child’s initial phenotype $\zeta_{/\theta} = 0$. Also assume that $\varepsilon_{i,j,c,t}^\theta = \eta_{j,c,t}^\theta = 0$, $\rho_{\theta} = 1$, $\rho_{l} = \pi_{i}$, $\rho_{Y} = 0$, $\pi_{Y} = 0$, and $w = 1$.

- In this case we get that the optimal investments are $I_{i,j,c,t}^* = I_{i',j,c,t}^* = I_{j,c,t}^*$:

$$I_{j,c,t}^* = \frac{\gamma_{Y}\rho_{l}}{1 - \gamma_{Y}(1 - 2\rho_{l})} \cdot Y_{i,j,a,t} \equiv \psi \cdot Y_{i,j,a,t} \quad (48)$$

- Note that this is the same solution as in Rustichini-et-al (2016).
The equilibrium law of motion also becomes

\[
\log(Y_{i,a,t+1}) = a_Y + \rho I \log(\psi) + \rho I \log(Y_{i,j,a,t}) \\
+ \pi \theta g_{i,j,t} + \pi \theta \varepsilon_{i,j,c,0,t} + \nu_{i,j,t+1}
\]  

(49)

- Note that while the absence of interaction effects implies that the IGE is constant, the initial phenotype (and therefore genes) affects the long-run level of income.
- The IGE coefficient is then given by

\[
\beta = \frac{\rho_I + 0.5\pi_\theta}{1 + 0.5\pi_\theta \rho_I}.
\]
Special Case: Becker-Tomes

\[
\log(Y_{i,a,t+1}) = a_Y + \rho I \log(\psi) + \\
+ (\rho I + 0.5\pi_\theta) \log(Y_{i,a,t}) - 0.5\rho I \pi_\theta \log(Y_{i,a,t-1}) \\
+ \nu_{i,j,t+1}^g + \epsilon_{i,j,c,0,t} + \epsilon_i^Y_{i,j,a,t+1} - 0.5\pi_\theta \epsilon_i^Y_{i,j,a,t} \quad (50)
\]
Denote the autocorrelations of child with father $r_{1,k}$, with grandfather $r_{3,k}$, with great-grandfather $r_{3,k}$, where $k = MZ, DZ$. And let

$$
\xi = \frac{(1 - \rho_I^2)(\sigma_{\nu,k}^2 + \sigma_{\epsilon}^2)}{(1 + 0.5\rho_I\pi_\theta)(\sigma_{\nu,k}^2 + \sigma_{\epsilon_0}^2) + (1 - 0.5\rho_I\pi_\theta)(\sigma_{\nu,k}^2 + \sigma_{\epsilon_Y}^2)}
$$

We can compute the autocorrelations

$$
\begin{align*}
r_{1,k} &= \rho_I + \pi_\theta \xi \\
r_{2,k} &= (\rho_I + \pi_\theta)r_{1,k} - \rho_I\pi_\theta \\
r_{3,k} &= (\rho_I + \pi_\theta)r_{2,k} - \rho_I\pi_\theta r_{1,k}
\end{align*}
$$

We can obtain identification using the genetic overlaps of twins.
Let

\[ \omega_{i,j,t} = \log(Y_{i,j,a,t+1}) \]
\[ A_{i,j,t} = g_{i,j,t-1} \]  \hspace{1cm} (51)
\[ C_{i,j,t} = \log(Y_{i,j,a,t}) \]
\[ E_{i,j,t} = \varepsilon_{i,j,a,t+1} \]

and

\[ \varepsilon_{A,i,j,t} = v_{i,j,t} \]
\[ \varepsilon_{E,i,j,t} = \varepsilon_{i,j,c,t} + \varepsilon_{i,j,a,t+1} \] \hspace{1cm} (52)
\[ \bar{\varepsilon}_{A,j,t} = v_{i',j,t} \]
\[ \bar{\varepsilon}_{C,j,t} = \rho_{\theta} \pi_{\theta} \varepsilon_{i,j,c,0,t} + \varepsilon_{i,j,c,t} + \varepsilon_{i,j,c,t+1} \]
Recall that \( X_{i,j,t} = (A_{i,j,t}, C_{i,j,t}, E_{i,j,t})' \), then our model is given by

\[
X_{i,j,t} = \Phi X_{i,j,t-1} + \varepsilon_{i,j,t},
\]

\[
\varepsilon_{i,j,t} = \Lambda \bar{\varepsilon}_j + \Psi \tilde{\varepsilon}_{i,j,t}.
\]

(53)

Note that

- given that \( C_{i,j,t} \) is the only observable variable, the above system can be viewed as a type of a factor-augmented structural vector autoregression (FAVAR),
- a measurement system can be used to measure skills in the spirit of Heckman and co-authors,
- genetic markers can be employed as instruments for the gene variable.
Then Identification of the SVAR entails the problem of identifying

$$\Sigma_{\epsilon,i} = \Lambda \Sigma \bar{\epsilon} \Lambda' + \Psi \Sigma \tilde{\epsilon}, \Psi'$$  \hspace{1cm} (54)

Let $u_{i,j,t} = (\bar{\epsilon}_{j,t}, \tilde{\epsilon}_{i,j,t})'$ and $S_i = \text{diag}\{\Lambda_i, \Psi\}$.

**Assumptions:**

(i) $E(u_{i,j,t}) = 0$ for all $i, j, t$;

(ii) The covariance matrix of the structural errors is finite and block diagonal with variances that can be normalized in arbitrary ways subject to equation (54).

(iii) $S_i^{-1}$ exists.
**Theorem**

*Under Assumptions (i)-(iii), the parameters $\Delta = (\Phi, \Sigma_u)$ are generically identified as a system.*

Our identification scheme relies not only on the standard contemporaneous second moments, namely, the covariances between the outcomes of identical twins and fraternal twins, that are employed in the classical ACE analysis but also, on the intertemporal second moments, namely, the covariance between parents’ income and children’s outcomes.
Variance Decomposition

Work-in-progress

- We provide forecast variance decompositions that capture the share of the forecast variance that can be attributed to particular shocks.

- We contrast these variances to provide comparisons of the changes in the variance covariance structure of outcomes when one or both sources of genetic variation are switched off.

- These differences measure the contribution of genes to inequality across individuals even though family environment is endogenous.

- This answers the Goldberger critique.
VARIANCE DECOMPOSITION

SET THE INDIVIDUAL SPECIFIC GENETIC SHOCKS TO ZERO

\[
\Sigma_{u,MZ} = \begin{pmatrix}
\rho^2 \pi^2_g & 0 & 0 \\
0 & \rho^2 + 2 & 0 \\
0 & 0 & \rho^2 + 2
\end{pmatrix}.
\]

\[
\Sigma_{u,DZ} = \begin{pmatrix}
0.25 \rho^2 \pi^2_g & 0 & 0 \\
0 & \rho^2 + 2 & 0 \\
0 & 0 & \rho^2 + 2
\end{pmatrix}.
\]

- The variance of income due to genes will be smaller than the shares of \(A\) and \(E\) if \(\rho < \sqrt{2/((\pi^2_g - 1))}\) in the case of MZ twins and if \(\rho < \sqrt{8/((\pi^2_g - 4))}\) in the case of DZ twins.
VARIANCE DECOMPOSITION

GENES ARE COMPLETELY DETERMINISTIC

\[ \Sigma_{u,DZ} = \begin{pmatrix}
0 & 0 & 0 \\
0 & \rho^2 + 2 & 0 \\
0 & 0 & \rho^2 + 2
\end{pmatrix}. \]
Variance Decomposition

• Contrast these variances to the variance covariance matrix as generated by the data, \( \Sigma \varepsilon_i \), to provide comparisons of the changes in the variance covariance structure of outcomes when one or both sources of genetic variation are switched off.

• These differences measure the contribution of genes to inequality across individuals even though family environment is endogenous.

• This provides an answer by overcoming Goldberger’s critique.
CONCLUSION

• We propose a dynamic ACE to identify the genetic and environmental effects in twin studies using a structural VAR approach based on a hybrid model of Heckman and Mosso (2014) with Becker-Tomes model as applied to twins.

• This shifts the focus from the regression coefficients of the ACE model to variance contributions of innovations in genes versus environment.

• We replace the statistical assumptions with behavioral assumptions.

• Next steps
  • Allow complementarities between investments and initial endowment, as well as interactions which will result in a nonlinear SVAR.

  • Calibration