On the Origins of Socio-Economic Inequalities: Evidence from Twin Families

Paul Bingley[a], Lorenzo Cappellari[b,d], Konstantinos Tatsiramos[c,d]

[a] VIVE Denmark
[b] Università Cattolica Milano
[c] University of Luxembourg
[d] Luxembourg Institute of Socio-Economic Research (LISER)

CRETE 2023 - Naxos

July 10-14, 2023

Supported by the Luxembourg National Research Fund (C21/SC/15771764)

Motivation

- Inherited **endowments** and family **environment** -> individual success
  - human capital, ability and skills (Becker and Tomes, 1979)

- Understanding **inequality** requires tracing individuals back to their **social origins**

- **Two approaches** to study the origins of observed differences in socio-economic outcomes:

  1. **Behavioral genetics model** (ACE) using twins or adoptees data

     Decompose outcome variance into **proportion** explained by
     (i) **genes**, (ii) **shared environment** and (iii) **non-shared environment**
     [Taubman, 1976; Behrman and Taubman, 1989; Björklund, Jäntti and Solon, 2005, Cesarini et al., 2009a,b; Barnea et al., 2010; Cronqvist and Siegel, 2015; Sacerdote, 2007; Fagereng, Mogstad and Rønning, 2021]

  2. **Regression-based methods** on adoptees data

     Focus on the **intergenerational elasticity** of economic outcomes
     [Sacerdote, 2002; Bjorklund, Lindahl and Plug, 2006; Lundborg, Plug and Rasmussen, 2022]

- **Different conclusions** regarding the **relative importance** of **endowments** at birth versus **family environment** in explaining overall inequality

This Paper

- Introduce a **new approach** based on a **twin family design** that links twins with their spouses and children

- Combine data from the **Danish Twin Registry** with **population registers** to reconstruct the **families of the twins**

- Two main **advantages** of the **twin family design** over the classical twin design (CTD):

    1. Decompose **cross-sectional inequality** without imposing the rather restrictive **assumptions** of the CTD

    2. Estimate the influence of **genes** on inequality both **within** the twin's generation but also on **intergenerational transmission**

- Direct answer to the long-standing question on how much **parent-child transmission** depends on factors already determined at birth

- Decomposing **intergenerational correlations** with **adoptees** requires **separation** of children from their parents

Assumptions of the CTD

1. No genetic **assortative mating** among spouses (mating is random)
2. No **genetic dominance** - gene variants add up but do not interact affecting outcomes
3. There are no **gene-environment interactions**
4. MZ and DZ twins share environments to the same extent (**equal environments assumption**)
5. There is no selection of genes into environments (no **gene-environment correlation**)

Biases when CTD assumptions are violated

- Assortative mating increases DZ twins' resemblance
  (**downward** biased genetic effects)
- Dominance reduces DZ twins' resemblance
  (**upward** biased genetic effects)
- Gene-environment interactions
  (sign of bias ambiguous)
- Differentially shared environments
  (**upward** biased genetic effects)

Preview of Results - Years of Schooling

- Classical Twin Design (CTD) - use MZ and DZ twin pairs only
  - **genetic factors**: account for a third of the variation (34%)
  - **shared environments**: account for a quarter of the variation (24%)
  - average heritability from literature; 40 % (Mogstad and Torsvik, 2022)
- CTD+spouses to test for **assortative mating**
  - modest positive correlation between the genes of the twins and their spouses
  - **genetic factors**: (41%)
  - **shared environments**: (18%)
- CTD+children to test for **dominance**
  - **No evidence** for genetic dominance
- Test for **gene-environment interactions** leveraging an educational expansion
  - school proximity varies in the parent generation differentially between twin families
  - **No evidence** of gene-environment interactions
- CTD + spouses + children (**twin family design**)
  - **Equal environments assumption rejected**; MZ twins share their environments much more than do DZ twins
  - **genetic factors**: (9%)
  - **shared environments**: (50%)

Preview of Results - Other Outcomes

- Decomposition of **cross-sectional inequality** in the **twin family design** for Earnings, Income, Wealth:
  - **genetic factors**: about 15%
  - **shared environments**: about 35%
- Decomposition of **intergenerational elasticities** - share of **family environment** explaining the correlation of outcomes between parents and children
  - Earnings: 50%
  - Income: 60%
  - Wealth 70%
  - Education 80%
- Main Take-Away: **Family environment** is more important than previously thought

The Canonical Twin Design

$$y_i = a_i + c_{f(i)} + e_i$$

- y_i denotes the **long-term outcome** of person i in deviations from the population mean
- a_i: additive **genetic factor** shared by family members
- $c_{f(i)}$: common **environmental factor** shared by family members
- $e_{(i)}$: **non-shared** environmental factor capturing idiosyncratic deviations within the family
- Information on variances and covariances of outcomes identify the variances of the three components under the set of maintained assumptions

Identification in the CTD

- The **total variance** of y_i is
$$var(y_i) = \sigma_a^2 + \sigma_c^2 + \sigma_e^2$$

- The **covariance** of outcomes between **MZ twins** i and i' is
$$cov(y_i, y_i')_{MZ} = \sigma_a^2 + \sigma_c^2$$

- The **covariance** of outcomes between **DZ twins** i and i' is
$$cov(y_i, y_i')_{DZ} = 0.5\sigma_a^2 + \sigma_c^2$$

- The **genetic variance** component can be estimated as:
$$\sigma_a^2 = 2(cov(y_i, y_i')_{MZ} - cov(y_i, y_i')_{DZ})$$

- **Heritability**: $\sigma_a^2/(\sigma_a^2 + \sigma_c^2 + \sigma_e^2)$

Data

- **Danish Twins Registry** combined with administrative **population registers**
- Sample of 80,205 individuals, 33,915 **twins**, 26,682 are twin **spouses** and 19,608 are **children of twins**
- Information on educational attainment, earnings, income and wealth

Empirical correlations - years of education

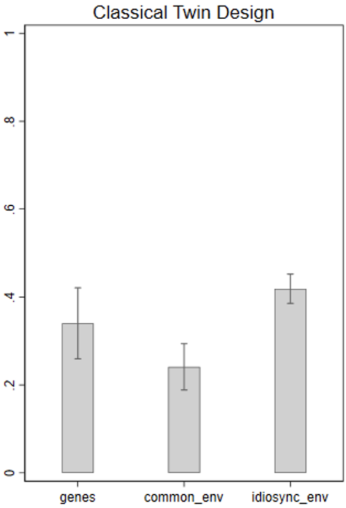
MZ twins (males)	0.58
MZ twins (females)	0.59
DZ twins (males)	0.45
DZ twins (females)	0.42
DZ twins (male-female)	0.36

Notes: The table reports empirical education correlations for different types of twin pairs.

Estimation

- We estimate parameters by Minimum Distance
- Match empirical variances and covariances to the corresponding moments generated by the model
- Equally Weighted Minimum Distance estimator (minimization weighted by the identity matrix)
- Adjust standard errors using the empirical matrix of fourth order moments

Decomposition of cross-sectional education inequality - CTD



Empirical correlations - years of education

Spouses	0.41
Father-Son	0.30
Mother-Daughter	0.34
Father-Daughter/Mother-Son	0.29
MZ Twin Uncle-Nephew	0.28
MZ Twin Aunt-Niece	0.26
MZ Twin Uncle (Aunt)-Niece (Nephew)	0.25
DZ Twin Uncle-Nephew	0.18
DZ Twin Aunt-Niece	0.20
DZ Twin Uncle (Aunt)-Niece (Nephew)	0.18
Male Cousins (MZ)	0.26
Female Cousins (MZ)	0.28
Male-Female Cousins (MZ)	0.25
Male Cousins (DZ)	0.17
Female Cousins (DZ)	0.13
Male-Female Cousins (DZ)	0.15
Brothers in-law (MZ-S)	0.28
Sisters in-law (MZ-B)	0.33
Brother-Sister in-law (MZ-BS)	0.33
Brothers in-law (DZ-S)	0.25
Sisters in-law (DZ-B)	0.29
Brother-Sister in-law (DZ-BS)	0.28

Note: The table reports empirical correlations for different family relations. MZ: monozygotic twins; DZ: dizygotic twins; MZ-S=MZ twins are sisters; MZ-B=MZ twins are brothers; MZ-BS=sister (brother) in-law is wife (husband) of MZ brother (sister); DZ-S=DZ twins are sisters or brother in-law is husband of DZ sister; DZ-B=DZ twins are brothers or sister in-law is wife of DZ brother; DZ-BS=sister (brother) in-law is wife (husband) of DZ brother (sister).

- The tendency for individuals to mate with partners that are similar to themselves
- **Extend** the CTD to include the **spouses of twins**
- The **covariance** of outcomes between **MZ twins** i and i' is

$$cov(y_i, y_i')_{MZ}^{T_1,T_2} = \sigma_a^2 + \sigma_{cT}^2$$

- The **covariance** of outcomes between **DZ twins** i and i' is

$$cov(y_i, y_i')_{DZ}^{T_1,T_2} = 0.5(1+\delta)\sigma_a^2 + \sigma_{cT}^2$$

  δ is the **assortative mating parameter**: the correlation of genotypes among spouses (see Bowles and Gintis, 2002)

- The **covariance** of outcomes between **spouses** is

$$cov(y_i, y_i')^{T_j S_j} = \delta\sigma_a^2 + \sigma_{cS}^2, \quad j = 1, 2$$

- For identification we need the **covariances** between **siblings-in-law**

Assortative mating - Siblings-in-law

- Siblings-in-law add **four** moment restrictions and **one** parameter
- The **covariance** of outcomes between **MZ twins** and the **spouse** of the **co-twin** is

$$cov(y_i, y_i')^{T_j,S_k}_{MZ} = \delta\sigma^2_a + \sigma^2_{cSL}, \quad j,k = 1,2 \quad j \neq k$$

- The **covariance** of outcomes between **DZ twins** and the **spouse** of the **co-twin** is

$$cov(y_i, y_i')^{T_j,S_k}_{DZ} = \delta(0.5(1+\delta))\sigma^2_a + \sigma^2_{cSL}$$

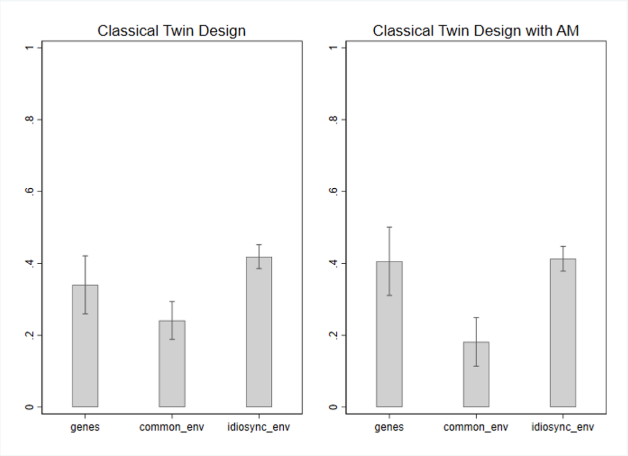
- The **covariance** of outcomes between the **spouses** of **MZ twins** is

$$cov(y_i, y_i')^{S_1S_2}_{MZ} = \delta^2\sigma^2_a + \sigma^2_{cS}$$

- The **covariance** of outcomes between the **spouses** of **DZ twins** is

$$cov(y_i, y_i')^{S_1S_2}_{DZ} = \delta^2(0.5(1+\delta))\sigma^2_a + \sigma^2_{cS}$$

Decomposition of cross-sectional education inequality - CTD vs. AM



Genetic dominance

- Refers to the relationship between the two versions of each gene inherited from each parent (25% prob for two siblings to inherit the same combination)

- **Extend** the CTD to include the **children of twins**

$$y_i = a_i + c_{f(i)} + d_i + e_i$$

- The total variance of y_i is

$$var(y_i) = \sigma_a^2 + \sigma_{cT}^2 + \sigma_d^2 + \sigma_e^2$$

- The **covariance** of outcomes between **MZ twins** i and i' is

$$cov(y_i, y_i')_{MZ}^{T_1,T_2} = \sigma_a^2 + \sigma_{cT}^2$$

- The **covariance** of outcomes between **DZ twins** i and i' is

$$cov(y_i, y_i')_{DZ}^{T_1,T_2} = 0.5\sigma_a^2 + \sigma_{cT}^2 + 0.25\sigma_d^2$$

- For identification we need the covariances between **uncle/aunt-nephew/niece** and between **cousins**

- They don't depend on dominance and genetic link depends on zygosity

Genetic dominance - additional moment restrictions

- The covariance of outcomes between an **MZ uncle/aunt** and a **nephew/niece** is:

$$cov(y_i, y_i')^{T_j, C_k}_{MZ} = 0.5\sigma_a^2 + \sigma_{cA}^2, \quad j,k = 1,2, \quad j \neq k$$

- The covariance of outcomes between a **DZ uncle/aunt** and a **nephew/niece** is:

$$cov(y_i, y_i')^{T_j, C_k}_{DZ} = 0.5^2\sigma_a^2 + \sigma_{cA}^2, \quad j,k = 1,2, \quad j \neq k$$

- The covariance of outcomes between cousins from **MZ twin parents** is:

$$cov(y_i, y_i')^{C_1, C_2}_{MZ} = 0.25\sigma_a^2 + \sigma_{cC}^2$$

- The covariance of outcomes between cousins from **DZ twin parents** is:

$$cov(y_i, y_i')^{C_1, C_2}_{DZ} = 0.50 * 0.25\sigma_a^2 + \sigma_{cC}^2$$

- Uncle-aunt/nephew-nice and cousins add **four** moment restrictions and **two** environmental parameters

Gene-environment interaction

- Genetic expressions mediated by the environment
- 1937 **school reform**: require municipalities to provide 8th and 9th grade teaching
- **Treated** twin pair (52%): **distance** between the parish of the **closest school** and the **parish of birth** was below median at age 14
- Estimate the CTD model **separately** for the two groups
- **Genetic** component is **not statistically different** between the two groups [$\chi^2(1) = 0.33$, p-value=0.566]

Differential shared environments

- **Extend** the CTD to include **spouses** and **children of twins** (twin family design)

- Hereogeneity of environmental sharing
  1. Zygosity of the twin pair
  2. Gender composition of the family tie

- The **covariance** of outcomes between **MZ twins** i and i' (where $g = 0, 2$) is

$$cov(y_i, y_i')^{T_1,T_2}_{gMZ} = \sigma^2_a + \sigma^2_{cTgMZ}$$

- The **covariance** of outcomes between **DZ twins** i and i' (where $g = 0, 1, 2$) is

$$cov(y_i, y_i')^{T_1,T_2}_{DZ} = 0.5\sigma^2_a + \sigma^2_{cTgDZ}$$

- For identification we need the **covariances** between **parent-child** and between **twins/co-parents** and their **niblings** (nephews/nieces)

- Key to **identification**: **spouses** of the twin pair only share the **environment** with the **child** of the **co-twin**

Differential shared environments - Additional moment restrictions

- The covariance of outcomes between **parent** and **child** is

$$cov(y_i, y_i')_g^{P_j, C_j} = 0.5\sigma_a^2 + \sigma_{cIg}^2, \quad P = S, T, \quad j = 1, 2$$

- The covariance of outcomes between **MZ twins** and their **niblings** is

$$cov(y_i, y_i')_{gMZ}^{T_j, C_k} = 0.5\sigma_a^2 + \sigma_{cAgMZ}^2, \quad j = 1, 2 \quad j \neq k$$

- The covariance of outcomes between **DZ twins** and their **niblings** is

$$cov(y_i, y_i')_{gDZ}^{T_j, C_k} = 0.5 * 0.5\sigma_a^2 + \sigma_{cAgDZ}^2, \quad j = 1, 2 \quad j \neq k$$

- The covariance of outcomes between **MZ twin spouses** and their **niblings** is

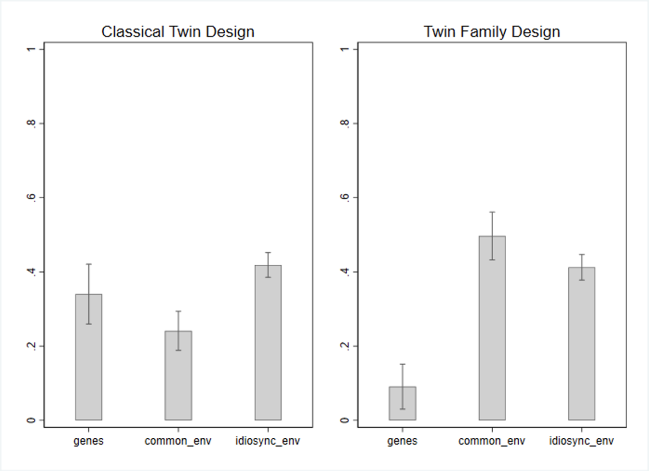
$$cov(y_i, y_i')_{gMZ}^{S_j, C_k} = \sigma_{cAgMZ}^2, \quad j = 1, 2 \quad j \neq k$$

- The covariance of outcomes between **DZ twin spouses** and their **niblings** is

$$cov(y_i, y_i')_{gDZ}^{S_j, C_k} = \sigma_{cAgDZ}^2, \quad j = 1, 2 \quad j \neq k$$

Decomposition of cross-sectional education inequality

CTD vs. Twin Family Design

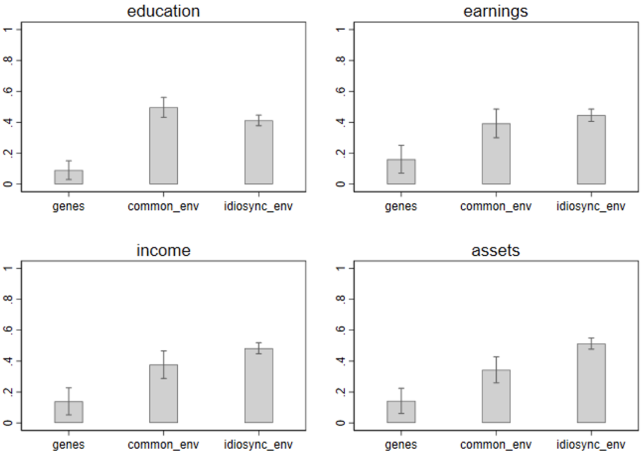


Other outcomes

- Annual pre-tax **labor earnings** from income tax returns
- **Disposable income**: personal gross income - taxes + transfers
- Wealth measured as taxable **assets** (financial and non-financial excluding liabilities)
- Permanent earnings, income, wealth: **average percentiles** in the 30-50 age window

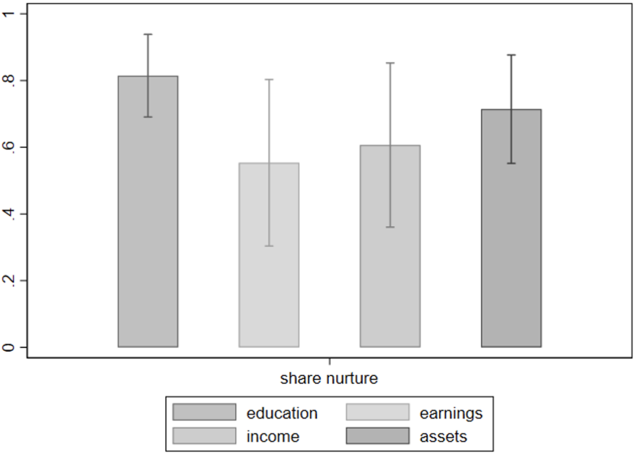
Decomposition of cross-sectional education inequality

All outcomes



Decomposition of intergenerational elasticities

All outcomes



Conclusion

- We quantify the relative importance of shared **genetic** and **environmental** factors in explaining the variance of economic outcomes
- We propose a **twin family model** linking twins with their children and co-parents
  - Extends the canonical twin design relaxing and testing for its restrictive **assumptions**
- Decompose both **cross-sectional inequality** and **intergenerational elasticities** (IGE)
  - Direct evidence on long-lasting question of role of factors determined at birth for **parent-child transmission**
  - IGE so far only decomposed using adoptees (separation of children from parents)
- **Differentially** shared environments are key
  - They explain half of the **variance** in **years of schooling**, while **genetic factors** explain only **nine percent**
  - We find similar shares for **earnings, income and wealth**
  - **Intergenerational transmission**: shared environments explain 50 percent for **earnings**, 60 percent for **income**, 70 percent for **wealth**, and 80 percent for years of **education**
- **Family environment** is more important than previously thought