Heterogeneity in Intergenerational Transmission of Education: Evidence from Norway

Aline Bütikofer   Lukáš Lafférs   Kjell Salvanes

NHH · UMB / MUNI / NHH · NHH

YEM Brno 2026

Motivation

The classical fact

Children of more-educated parents tend to get more education themselves.

> Usually a **single coefficient**.
> For millions of people.

Our question

Heterogeneity

- Link the same for everyone?
- Or does it vary ?
- Where it does *not* show up?

Approach: *let the data speak for itself.* A causal forest approach.

literature four strands

Data

Norwegian registry, 3 generations

- child, parents, grandparents
- education thresholds (mand., HS, Bc, MSc)
- cohorts 1955–1980; 46 regions
- cognitive-ability decile at conscription
- parental earnings ranks; grandfathers' edu

This draft

- Bc thres.: W = father Bc, Y = child Bc
- $n = 100{,}000$ (full $\sim$449k forthcoming)
- 99% male (ability via conscription)
- $\Pr(W) = 16.7\%$, $\Pr(Y) = 29.6\%$

What we estimate

$$\tau(x) = E[\,Y \mid W = 1, X = x\,] - E[\,Y \mid W = 0, X = x\,]$$

Δ in the prob. that the child holds a Bc, between sons with father with Bc. vs not, cond. on X.

X contains

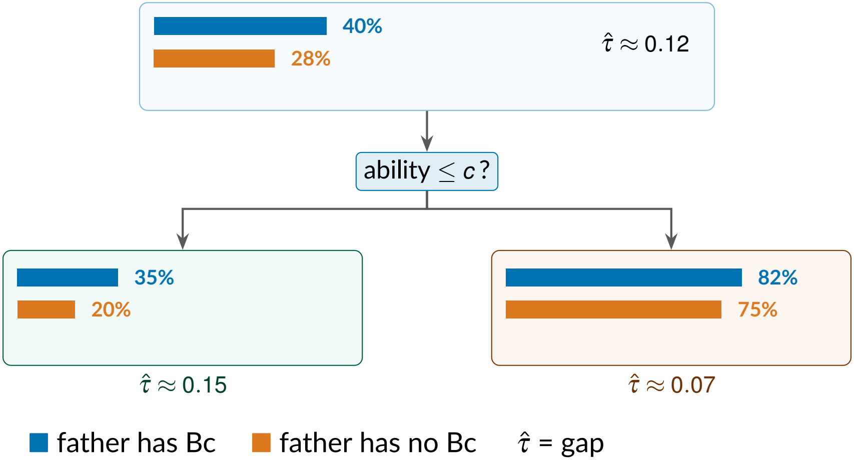
ability, cohort, region, mother edu, parental earnings, both grandfathers' edu.

Estimator

Causal forest (Athey, Tibshirani, Wager 2019); honest trees; clustered SEs by region.

Not a causal effect – a conditional association.

visualization cf details causal language


40%
28%
$\hat{\tau} \approx 0.12$
ability $\leq c$?
35%
20%
82%
75%
$\hat{\tau} \approx 0.15$
$\hat{\tau} \approx 0.07$
father has Bc
father has no Bc
$\hat{\tau}$ = gap

Headline: large average, real heterogeneity

Average overlap-weighted CATE

$$\hat{\tau} = 0.124$$

(SE $= 0.004$, clustered by region)

Sons whose fathers have a Bc are about **12 pp** more likely to have a Bc, cond. on X.

Calibration test

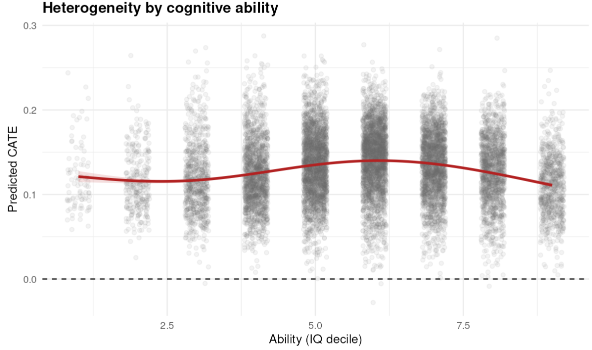
	Coef.	t
MFP (avg. pred.)	1.00	20.9
DFP (differential)	0.46	4.0

MFP ≈ 1**:** average well calibrated.

DFP ≈ 0.5, $t = 4$**:** real heterogeneity.

overlap RATE var. importance

Two moderators: ability (hump) + father earnings



BLP (debiased, joint)

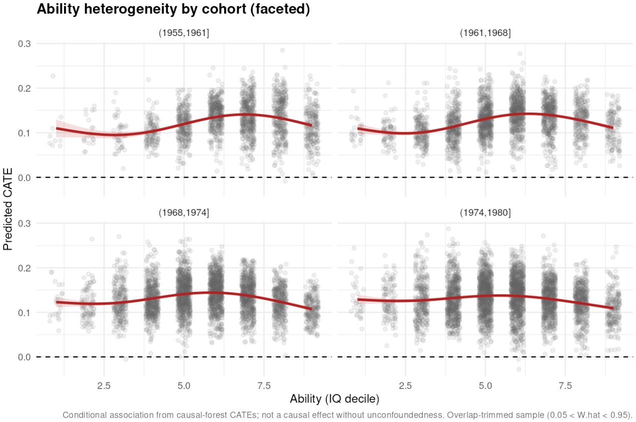
	Coef.	p
ability^2	-0.0079	0.001
father earn	0.0022	0.02
father earn^2	-5×10^{-5}	0.09

Ability: inverted U, peak in middle deciles – the headline moderator.

Father earnings: positive linear slope – a *weaker* secondary moderator. Concave curvature marginal.

by cohort ability table earnings

The hump is flattening – democratization?



- Hump shows up everywhere.
- Curvature **attenuates** later.
- Ability gradient **weakens over time.**

> Consistent with **democratization**: a child's Bc is less and less determined by ability.

by cohort | evolves | cohort levels

Where we do *not* find heterogeneity

BLP, other moderators

	Coef.	p
mother education	9.0×10^{-3}	0.35
mother earnings	1.8×10^{-5}	0.98
cohort	3.5×10^{-3}	0.14
pat. grandfather	-7.1×10^{-3}	0.13
mat. grandfather	-4.1×10^{-3}	0.60

(father earnings on previous slide – it *does* moderate, weaker than ability)

A result, not a caveat

- **Mother's edu**: no moderation – against assortative-mating.
- **Mother's earnings**: flat – contrast with father earnings.
- **Grandfathers**: no moderation in slope.
- **Cohort**: no moderation in level.

grandfather mother edu cohort flat

ability $\times$ earn earn $\times$ ability 3-way

Robust across alternative treatments

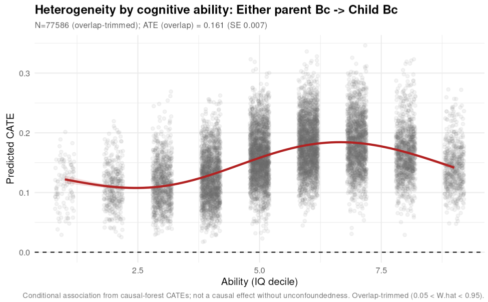
Four alternative W.

W definition	ATE	Hump?
Father Bc (main)	0.12	yes
Mother Bc	0.16	yes
Either parent Bc	0.16	yes
Both parents Bc	0.20	yes
Father MSc	0.14	**no**

Inverted U for every Bc treatment; flat for MSc (worst overlap, $\sim 75\%$ trimmed).

> The hump is a property of **Bc transmission**, not of which parent. Most able sons gain little regardless.

diagnostics all four plots



Either-parent-Bc (cleanest overlap).

Interpreting the hump

Story 1: threshold mechanics

bottom: Bc unreachable whatever the father does

top: Bc reachable whatever the father does

middle: **father's education tips the balance**

> **Our data cannot tell these apart.**

Story 2: differential investment

- educated fathers invest more when child shows promise
- ability and effort reinforce each other
- information / aspirations channels

Takeaways

- ***Agnostic, mechanical*** search.
- Father-to-son Bc transmission $\approx$ 12 pp on average.
- Real, significant **heterogeneity** – narrowly located.
- Where it *is*: ability (**inverted U**, headline) and **father earnings** (positive, weaker).
- Where it is *not*: mother's edu, mother's earnings, grandfathers, region, cohort level.
- Level **flat across cohorts**; the **ability hump flattens over time** – consistent with democratization.
- **Robust** across Bc treatments (mother / either / both); MSc shows no hump.

limitations next steps

Thank you.

Comments & questions welcome.

Appendix

Where this sits in the literature

What the literature has done

Average link: Black & Devereux (2011); Holmlund, Lindahl & Plug (2011); Hertz et al. (2007).

Causal ID: Black, Devereux & Salvanes (2005); Björklund, Lindahl & Plug (2006); Pronzato (2012).

Three generations: Lindahl, Palme, Sandgren Massih & Sjögren (2015).

Ability heterogeneity: Nordin & Rooth (2011); Lundborg, Nilsson & Rooth (2014).

full bibliography

What is missing

- prior work assumes a **parametric form**
- picks the moderator *in advance*
- no open-ended ML search
- silent on where heterogeneity is *absent*

> **Closest precedent:** Nordin & Rooth (2011) find a hump on Swedish data; we recover it in Norway with a *mechanical* forest.

back

Literature – four strands (detail)

- **Average link.** Black & Devereux (2011) and Holmlund, Lindahl & Plug (2011) survey decades of evidence; Hertz et al. (2007) put the global parent–child schooling correlation at ≈ 0.4 across 42 countries.
- **Causal identification.** Black, Devereux & Salvanes (2005) use a Norwegian compulsory-schooling reform and find the causal effect of a father's schooling close to zero. Björklund, Lindahl & Plug (2006) separate pre- and post-birth channels using Swedish adoptees. Pronzato (2012) revisits the Norwegian IV. Most of the OLS link reflects selection.
- **Three generations.** Lindahl, Palme, Sandgren Massih & Sjögren (2015) link four Swedish generations: persistence is stronger than an AR(1) would predict, which supports including grandparents as controls.
- **Ability heterogeneity.** Nordin & Rooth (2011) and Lundborg, Nilsson & Rooth (2014) use Swedish conscription tests and show the parent–child link varies with the son's cognitive ability – hump-shaped, peaking in the middle. Ability is grouped into pre-set bins.

full bibliography ◂ back

Bibliography

Intergenerational education

Black, S. & Devereux, P. (2011). Recent Developments in Intergenerational Mobility. *Handbook of Labor Econ.* 4B, Ch. 16.

Black, S., Devereux, P. & Salvanes, K. (2005). Why the Apple Doesn't Fall Far. *AER* 95(1), 437–449.

Björklund, A., Lindahl, M. & Plug, E. (2006). The origins of intergenerational associations: Lessons from Swedish adoption data. *QJE* 121(3), 999–1028.

Hertz, T. et al. (2007). The inheritance of educational inequality: International comparisons and fifty-year trends. *B.E. JEAP* 7(2), Art. 10.

Holmlund, H., Lindahl, M. & Plug, E. (2011). The causal effect of parents' schooling on children's schooling: A comparison of estimation methods. *JEL* 49(3), 615–651.

Pronzato, C. (2012). An examination of paternal and maternal intergenerational transmission of schooling. *J. Pop. Econ.* 25(2), 591–608.

Lindahl, M., Palme, M., Sandgren Massih, S. & Sjögren, A. (2015). Long-term intergenerational persistence of human capital: an empirical analysis of four generations. *J. Human Res.* 50(1), 1–33.

Ability heterogeneity

Lundborg, P., Nilsson, A. & Rooth, D.-O. (2014). Parental education and offspring outcomes: evidence from the Swedish compulsory School Reform. *AEJ: Applied* 6(1), 253–278.

Nordin, M. & Rooth, D.-O. (2011). Ability Heterogeneity in Intergenerational Mobility. Lund Univ. WP 2011:18.

Causal forest / ML inference

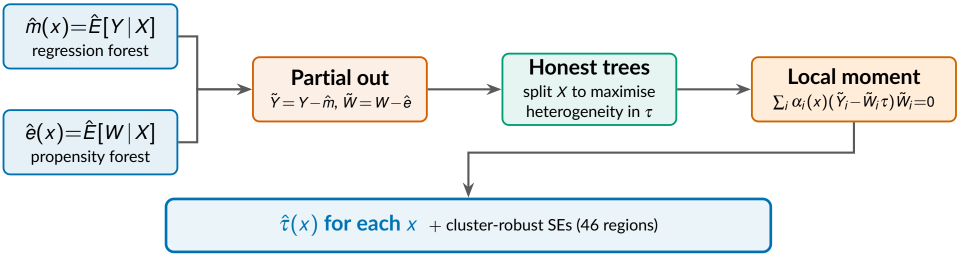
Athey, S., Tibshirani, J. & Wager, S. (2019). Generalized Random Forests. *Annals of Statistics* 47(2), 1148–1178.

Wager, S. & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *JASA* 113(523), 1228–1242.

Chernozhukov, V., Demirer, M., Duflo, E. & Fernández-Val, I. (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. NBER WP 24678 / arXiv:1712.04802.

◂ back

Causal forest – the picture



Mechanical pipeline: no parametric guess about how τ varies with X.

Honesty: each tree's splits are fit on a different subsample from the one used to estimate τ in the leaves – gives asymptotic normality (Wager & Athey, 2018).

◂ back

Causal forest in one slide

Athey, Tibshirani, Wager (2019)

ensemble of **honest** trees splits on X to maximise heterogeneity in τ

forest-weighted local moment at each x

Robinson / R-learner: subtract two nuisance forests

$$\hat{m}(x) = \hat{E}[Y \mid X{=}x], \;\; \hat{e}(x) = \hat{E}[W \mid X{=}x]$$

region-clustered SEs (46 clusters)

Reading guide

- individual $\hat{\tau}(x)$ noisy
- averages, rankings precise
- BLP for moderation

> **Diagnostics:** overlap, calibration, RATE.

back

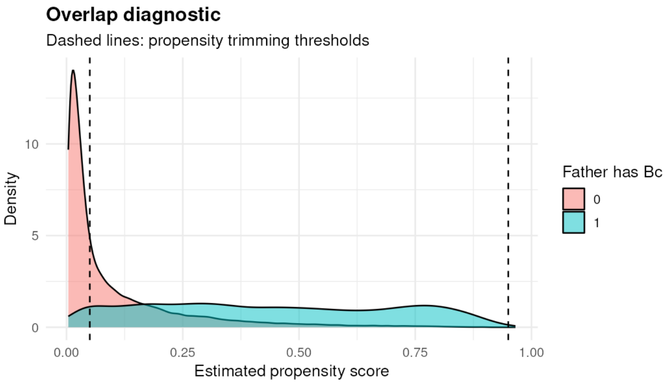
Causal language?

> A father's education is *not* randomly assigned. $\widehat{\tau}$ is a **conditional association**, not a causal effect.

- Unconfoundedness is implausible at the individual level (genes, home environment, schooling neighbourhood).
- We follow Black, Devereux & Salvanes (2005), Björklund, Lindahl & Plug (2006), Pronzato (2012) in reading the OLS gradient as mostly selection.
- Our contribution is the *shape* of the association across the covariate space, not its causal magnitude.

back

Overlap diagnostic



Density of $\hat{e}(X)$ by treatment status. Dashed lines: trimming thresholds. About 47% of obs trimmed – mostly at low propensities.

back

RATE (AUTOC, overlap)

$$\text{RATE} = 0.012 \quad \text{(SE 0.007)}$$

- Modest **targeting** signal.
- Most heterogeneity sits in observable dimensions (especially ability).
- Individual CATEs are too noisy to rank reliably – use group averages.

back

Variable importance

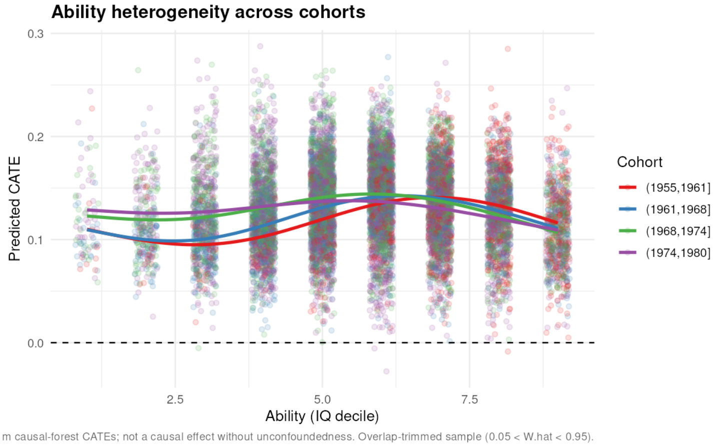
Variable	Importance
mother edu	0.302
father earn pct.	0.261
ability	0.184
mother earn pct.	0.078
mat. grandfather	0.059
cohort	0.051
pat. grandfather	0.038
region	0.026

Variable importance = how often each variable is split on when the causal forest looks for $\tau(x)$ heterogeneity. Identifies where moderation lives, but is silent on sign and shape, and inflates for variables correlated with true moderators.

The BLP is the right object for moderation.

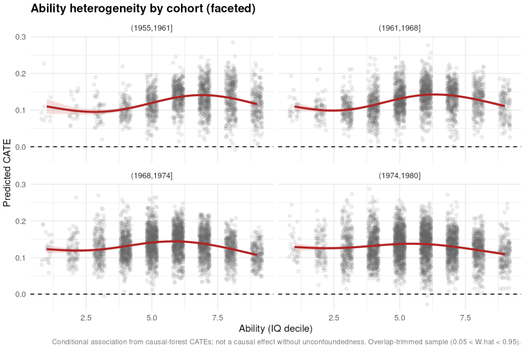
back

Ability by cohort – coloured overlay



◂ back

...and the shape evolves in time



> The hump shows up in every cohort group. It flattens a little in later cohorts – the ability gradient weakens over time.

back

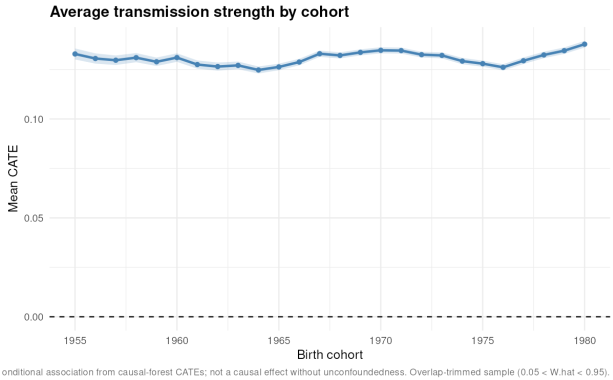
Same story, subgroup ATEs

Ability decile	ATE	SE	n
1	−0.025	0.021	1,935
2	0.070	0.020	4,960
3	0.089	0.017	10,273
4	0.105	0.012	17,677
5	0.143	0.009	21,664
6	0.153	0.008	18,854
7	0.123	0.012	13,090
8	0.111	0.012	7,737
9	0.074	0.016	3,810

Gap between peak (decile 6) and bottom (decile 1) **larger than the full ATE**. Bottom decile indistinguishable from zero.

back

Cohort: level is flat…



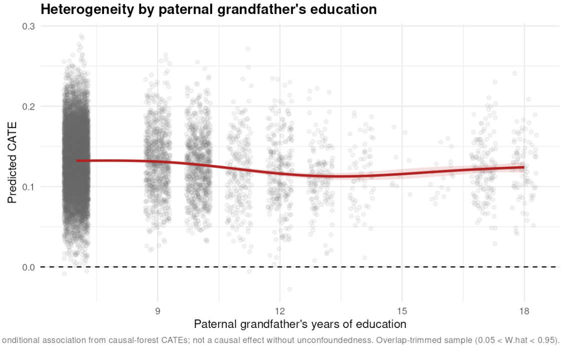
Subgroup ATEs

Cohort bin	ATE
(1955, 1961]	0.126
(1961, 1968]	0.119
(1968, 1974]	0.130
(1974, 1980]	0.119

Average transmission has been **stable** over the period.

back

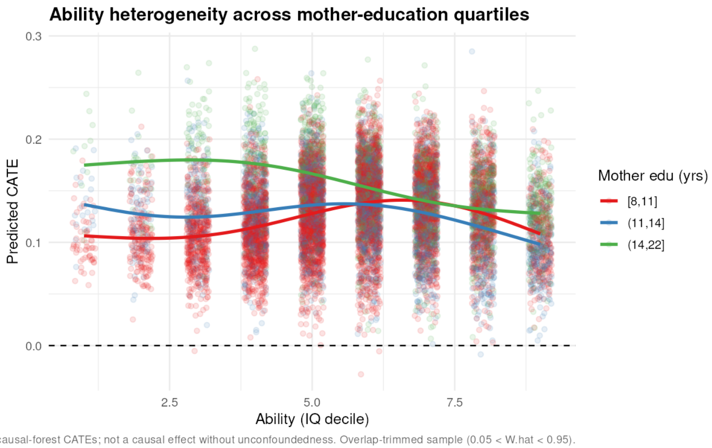
Paternal grandfather’s education



The slope is negative – partial substitution across generations. When the grandfather was more educated, the father-to-son link is slightly weaker.

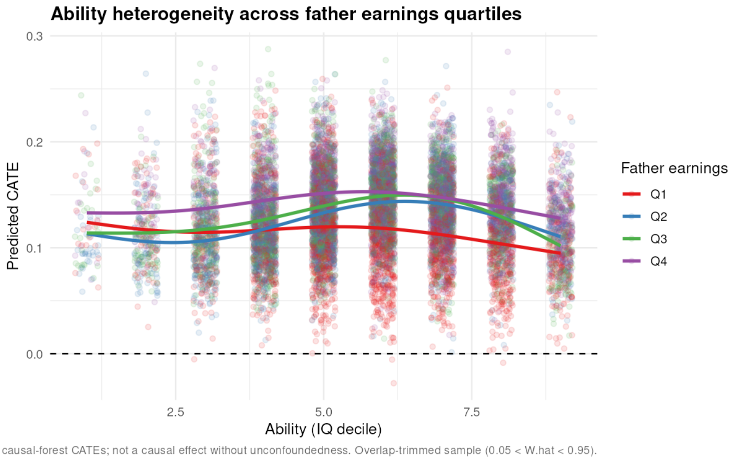
back

Ability by mother education quartile



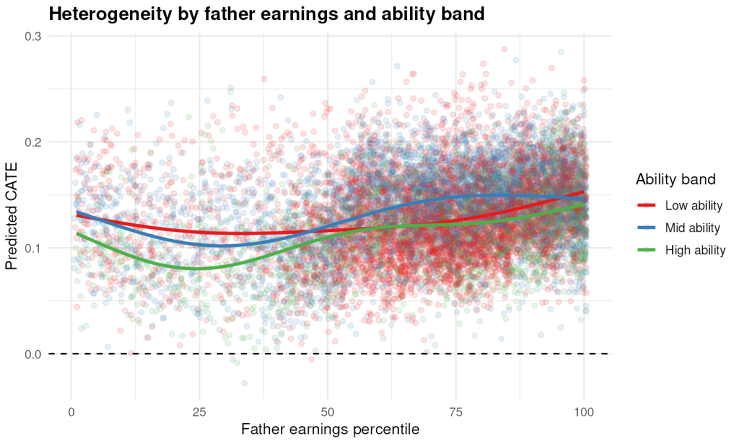
back

Ability by father earnings quartile



back

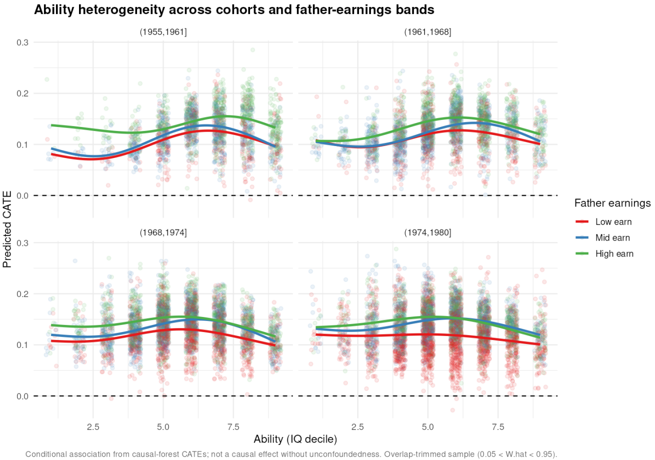
Father earnings, by ability band



om causal-forest CATEs; not a causal effect without unconfoundedness. Overlap-trimmed sample (0.05 < W.hat < 0.95).

back

Three-way: ability × cohort × earnings



back

Alternative treatments – diagnostics

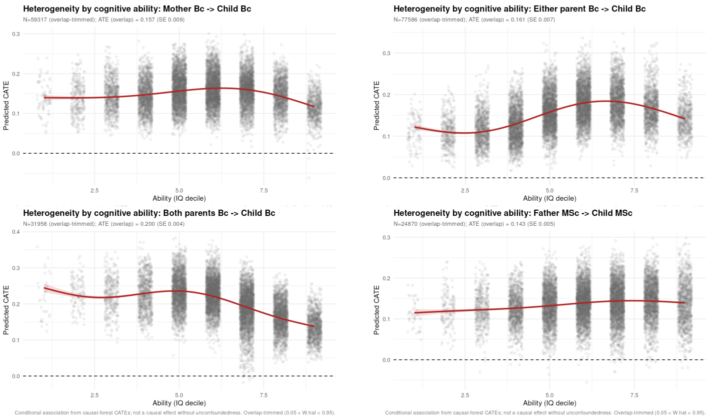
W definition	$\Pr(W{=}1)$	ATE	SE	% trimmed	DFP (t)
Father Bc (main)	0.167	0.124	0.004	47%	0.46 (4.0)
Mother Bc	0.125	0.157	0.009	41%	0.57 (5.2)
Both parents Bc	0.070	0.200	0.004	68%	0.75 (8.5)
Either parent Bc	0.222	0.161	0.007	22%	0.68 (8.3)
Father MSc	0.065	0.143	0.005	75%	0.14 (1.3)

MFP ≈ 1 in every spec. DFP significant for all Bc treatments, *not* for MSc. Either-parent-Bc has the best overlap; MSc the worst.

MSc mother both either

back

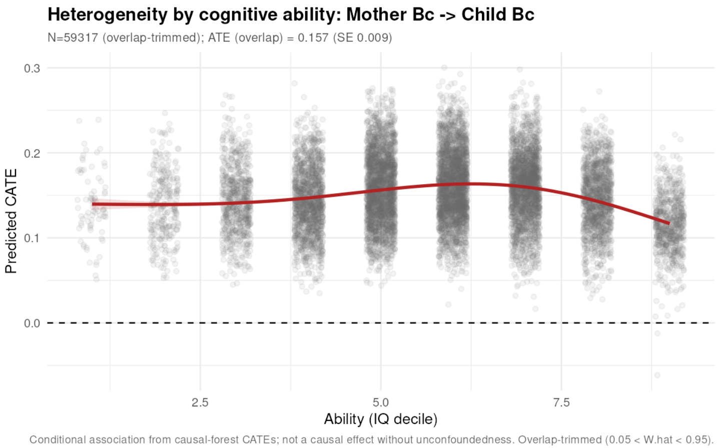
Alternative treatments – ability × CATE



Mother Bc (top-left), either parent Bc (top-right), both parents Bc (bottom-left), father MSc (bottom-right).

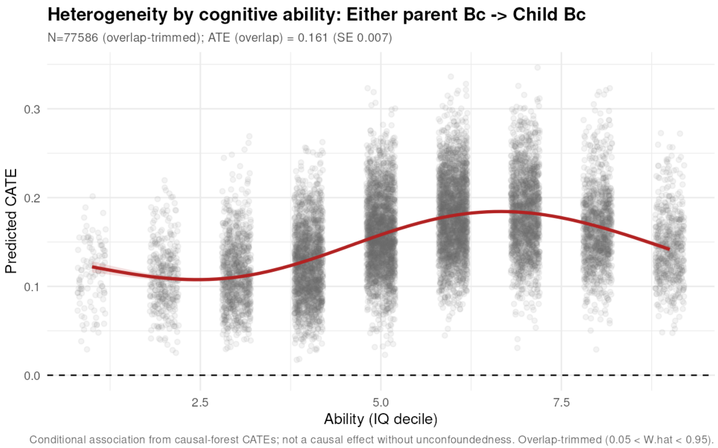
back

Mother Bc → Child Bc



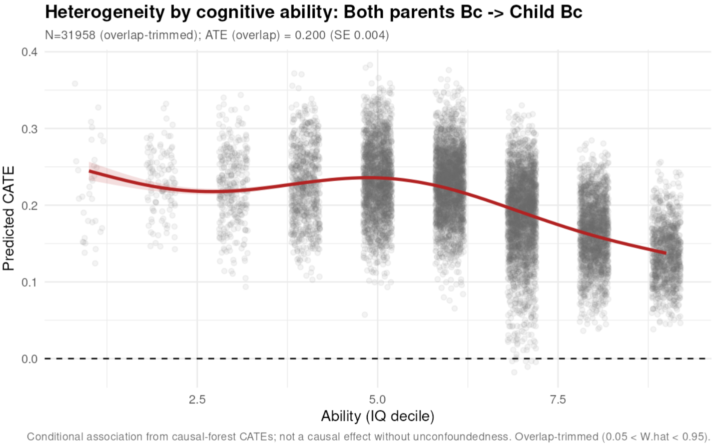
back

Either parent Bc $\rightarrow$ Child Bc



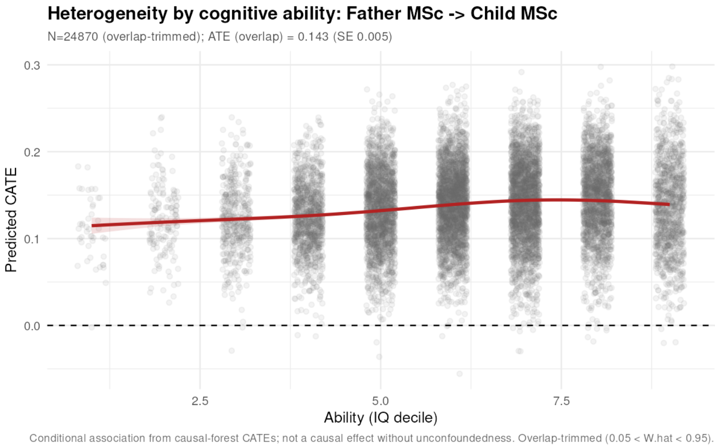
back

Both parents Bc $\longrightarrow$ Child Bc



◂ back

Father MSc → Child MSc



Flat / mildly rising; no hump. Heterogeneity not detected (DFP $t=1.3$), 75% trimmed – read with care.

back

Limitations

- **Causal language:** estimates are conditional associations – a father's education is not exogenously assigned.
- **Sample:** conditioning on ability restricts us to father–son pairs. Daughters (no ability score) on the to-do list.
- **Forest noise:** individual CATEs spread about twice as wide as they should be. Use ranks and averages, not individual predictions.
- **Power:** 100k subsample for this draft; secondary moderators need the full $\sim$449k.

back

Next steps

- full-sample replication ($\sim$449k)
- repeat at the master's-degree threshold
- daughter-only specification (no ability score)
- exogenous variation: compulsory-schooling reforms, regional expansions of higher education
- richer subgroup tables and figures for the paper

back